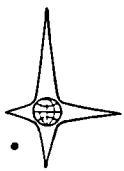
БИБЛИОТЕКА СБОРНИКА
МАТЕМАТИКА

Р. Галлагер

Коды с малой плотностью проверок на четность



ИЗДАТЕЛЬСТВО
« М И Р »

LOW-DENSITY PARITY-CHECK CODES

by

Robert G. Gallager

M. I. T. PRESS, CAMBRIDGE,
MASSACHUSETTS
1963

БИБЛИОТЕКА СБОРНИКА „МАТЕМАТИКА"

Р. Дж. ГАЛЛАГЕР

КОДЫ С МАЛОЙ ПЛОТНОСТЬЮ ПРОВЕРОК НА ЧЕТНОСТЬ

Перевод с английского
Л. ШЕВЕРДЯЕВА

Под редакцией
Р. Л. ДОБРУШИНА

ИЗДАТЕЛЬСТВО «МИР»
Москва 1966

У.Д.К. 621.391.015

Книга является монографией известного американского специалиста в области теории информации. Она посвящена практически важному классу алгебраических кодов и разработке легко реализуемых методов кодирования и декодирования для передачи информации по реальным каналам связи со скоростью, приближающейся к пропускной способности канала.

Книга предназначена для научных работников и инженеров, занимающихся теорией информации и теорией кодирования, а также для математиков, интересующихся приложениями, и военных специалистов. Она доступна аспирантам и студентам старших курсов университетов, энергетических институтов и институтов связи.

Редакция литературы по математическим наукам

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Предлагаемая вниманию читателя книга Галлагера является уже третьей в серии переведенных на русский язык небольших монографий по теории кодирования. Первая из этих монографий — книга Возенкрафта и Рейффена «Последовательное декодирование», вышедшая в русском переводе два года назад, хорошо известна читателям-специалистам. Вторая — книга Месси «Пороговое декодирование» — недавно издана в русском переводе. Объединяет эти три книги многое. Все они написаны молодыми учеными, принадлежащими к ведущей в США научной школе в области теории информации — школе, работающей в Массачусетском технологическом институте, и все они представляют собой изложение оригинальных исследований их авторов, предложивших различные подходы к решению важнейшей проблемы техники связи: проблемы построения практически осуществимых методов кодирования и декодирования, позволяющих вести надежную передачу информации по реальным каналам связи со скоростью, приближающейся к теоретической границе — пропускной способности канала. Чтобы полностью исчерпать список основных идей в этой проблематике, нужно добавить к темам этих трех перечисленных книг, пожалуй, лишь широко известный метод циклических кодов, хорошо изложенный в книге Питерсона «Коды, исправляющие ошибки». Кроме того, стóит специально отметить небольшую, но важную статью нашего соотечественника М. Пинскера, опубликованную в журнале «Проблемы передачи информации» (№ 1 за 1965 г.), которая развеивает предрассудок (разделяемый, по-видимому, специалистами Массачусетской школы теории информации), состоящий в том, что возможность передачи информации со сколь угодно малой вероятностью ошибки и без чрезмерно большой по объему вычислительной работы при декодировании всегда ограничена

некоторой скоростью, меньшей пропускной способности канала.

Метод Галлагера основан на простой, но остроумной математической идее. Он рассматривает групповой код, проверочная матрица которого в основном состоит из нулей и содержит лишь небольшое число единиц. Простота структуры проверочной матрицы создает возможность предложить простые алгоритмы декодирования. Желание доказать, что введенный алгоритм декодирования дает ошибку лишь с малой вероятностью, приводит автора к необходимости провести сложные математические построения как комбинаторного, так и аналитического характера. Один из введенных при этом методов — оценка вероятности ошибки декодирования через структуру кодовых расстояний — интересен для теории кодирования и вне рамок основной темы книги. Тем не менее автору не удается создать достаточно полную теорию кодов с малой плотностью проверок на четность; то, что им получено, — это скорее фрагменты такой теории. Конечно, установленные в книге теоретические результаты и приведенные в ней результаты моделирования на вычислительных машинах создают должную эмпирическую уверенность в пригодности алгоритмов декодирования; однако, вступая в противоречие с замечанием, мельком сделанным автором в книге, автор предисловия думает, что дальнейшие теоретические исследования вопроса были бы плодотворными как с математической, так и с инженерной точки зрения.

Книга написана четко и ясно, но сжато. Предполагается предварительное знакомство с элементами теории групповых кодов (например, достаточно ознакомления с первыми главами упомянутой выше книги Питерсона). Кое-где используются также элементы общей теории информации (с которыми можно ознакомиться, например, по книге Фано «Передача информации»). Книга Галлагера доступна и интересна быстро растущему кругу инженеров и математиков, работающих в области теории кодирования и ее приложений.

Р. Л. Добрушин

ПРЕДИСЛОВИЕ АВТОРА

Теорема кодирования для канала с шумами, установленная К. Э. Шенноном в 1948 г., показала инженерам-связистам возможность получения сколь угодно малых частот ошибок без потерь в скорости передачи информации. Главными препятствиями в практическом использовании этой теоремы были сложность оборудования и большое время вычислений, требуемых для декодирования принятой при наличии шума информации.

В настоящей монографии приводится методика получения высоких скоростей передачи и сколь угодно малых вероятностей ошибки с использованием разумного объема оборудования. Ответ на вопрос о преимуществах и недостатках этой методики в сравнении с другими ни в коем случае не прост и не однозначен: он зависит прежде всего от типа канала и от требуемых результатов. Однако надежда на то, что развитые здесь концепции приведут к созданию новых и лучших методов кодирования важнее, чем конкретная методика.

Главы монографии построены так, что каждую можно читать независимо (за исключением гл. 5). В гл. 1 обсуждается общее состояние предмета, суммируются результаты и кратко сравнивается кодирование с малой плотностью проверок на четность с другими методами кодирования. В гл. 2 исследуются расстояния между кодовыми словами в кодах с малой плотностью проверок, а в гл. 3 эти результаты используются при оценке вероятности ошибки декодирования, которой можно достичь для таких кодов в широком классе каналов с двоичным входом. Результаты гл. 3 можно применить непосредственно к любому коду или классу кодов, для которых можно найти оценку свойств расстояния. В гл. 4 приводится простой алгоритм декодирования в кодах с малой плотностью проверок и для него исследуется вероятность

ошибки декодирования. В гл. 5 предыдущие результаты кратко обобщаются на недвоичные каналы, а в гл. 6 приводятся результаты моделирования алгоритма декодирования с малой плотностью проверок на четность на вычислительной машине.

Эта книга представляет собой расширенный и пересмотренный вариант моей докторской диссертации, законченной в 1960 г. в Электротехническом отделе МТИ. Я признателен моему руководителю, проф. Питеру Элайесу, и моим оппонентам, проф. Роберту М. Фано и Джону М. Возенкрафту, за помощь и ободрение как во время подготовки диссертации, так и после этого.

Это исследование стало возможным отчасти благодаря поддержке, оказанной Исследовательской лабораторией электроники Массачусетского технологического института, которая частично поддерживается армией США, Отделом научных исследований военно-воздушных сил и Отделом военно-морских исследований; дополнительная поддержка была получена от Национального научного фонда и от Национального института здоровья.

Большая часть гл. 4 перепечатана с разрешения редакторов из статьи автора в Transactions of the I. R. E., IT-9, стр. 21—28.

Экспериментальные результаты гл. 6 были получены частично при поддержке Авиационного исследовательского центра в Риме и Вычислительного центра МТИ.

Роберт Дж. Галлагер

Кембридж, Массачусетс
Июль 1963 г.

Глава 1

ВВЕДЕНИЕ

1.1. Кодирование при передаче цифровой информации

За последние годы резко возросла потребность в эффективных и надежных системах передачи информации. Это вызвано целым рядом причин и среди них растущим применением аппаратуры автоматической обработки информации и увеличивающейся потребностью в связи на большие расстояния. Попытки создать информационные системы с использованием обычных методов модуляции и телефонной техники

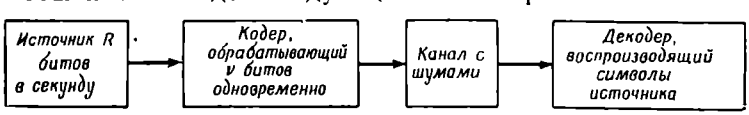


Рис. 1.1. Блок-схема системы связи.

приводили, как правило, к системам со сравнительно низкими скоростями передачи информации и высокой вероятностью ошибки.

Более глубокий подход к проблеме эффективности и надежности систем связи содержится в теореме кодирования для канала с шумами, доказанной К. Е. Шенноном в 1948 г. [4, 15]. Для того чтобы лучше понять смысл этой теоремы, рассмотрим рис. 1.1. Источник создает двоичные символы с некоторой фиксированной скоростью R. Кодер — устройство, обрабатывающее поступающую информацию, осуществляющее модуляцию и вообще делающее все необходимое для подготовки информации к передаче по каналу. Предположим, что кодер разбивает последовательность символов, создаваемых источником, на блоки по ν символов и обрабатывает одновременно символы

только одного блока. Символы на выходе кодера передаются по каналу и изменяются под действием каких-либо случайных возмущений или шума. Декодер обрабатывает символы на выходе канала и воссоздает с задержкой некоторый вариант символов, созданных источником. Теорема кодирования утверждает: для широкого класса моделей каналов существуют такие кодер и декодер, для которых вероятность P_e того, что декодер воспроизводит символ источника ошибочно, оценивается следующим образом:

$$e^{-\nu[E_L(R_t)+o(\nu)]} \leqslant P_e \leqslant e^{-\nu E(R_t)}.$$

Функции $E(R_t)$ и $E_L(R_t)$ зависят от канала, но не зависят от ν; они положительны при $R_t=0$, убывают с ростом R_t и обращаются в нуль при некоторой скорости C_t, называемой пропускной способностью канала. Нам нет необходимости рассматривать здесь точный вид этих функций и класс каналов, для которых справедлива теорема. Важно то, что длина кодовых ограничений ν — это основной параметр системы связи. Если мы хотим использовать канал эффективно, т. е. со скоростью R_t, близкой к пропускной способности C_t, и получить при этом удовлетворительную вероятность ошибки, значение ν следует выбрать достаточно большим.

На такую теорему реакция инженера очевидна: «Великолепно, но как построить такие кодеры и декодеры, если ν велико?» Соображения, показывающие, что требуемый для их построения объем памяти пропорционален 2^ν в том случае, когда кодер содержит в памяти сигнал, или кодовое слово, для каждого из возможных блоков по ν символов, действуют в достаточной мере отрезвляюще. К счастью, Элайес [3] и Рейффен [14] показали, что для широкого класса моделей каналов результаты, даваемые теоремой кодирования для канала с шумами, могут быть достигнуты при небольшой сложности кодера с помощью кодов с проверками на четность. Ниже мы подробнее рассмотрим этот факт.

Проблема простого, но эффективного декодирования, к сожалению, оказывается при больших ν гораз-

до более трудной, чем проблема кодирования. Было предложено достаточно большое число подходов к этой проблеме, для того чтобы доказать инженерам ценность теоремы кодирования. Эти подходы, однако, не были разработаны настолько, чтобы создание эффективной и надежной системы передачи информации стало вопросом повседневной техники.

Данная монография содержит детальное исследование одного из трех или четырех наиболее перспективных подходов к проблеме простого декодирования кодов с большой длиной кодовых ограничений. Цель опубликования этой работы состоит прежде всего в том, чтобы показать, как и где можно использовать такой метод кодирования и декодирования, найденный в результате этого подхода. Я надеюсь, кроме того, что она будет способствовать дальнейшему развертыванию исследований в этой области. Дальнейшие математические исследования, вероятно, не будут плодотворными, однако существует целый ряд интересных модификаций метода, которые можно было бы развить; нужно также провести большую экспериментальную работу.

Для того чтобы математически доказать некоторые факты, мы примем, что эти коды с малой плотностью проверок на четность будут использоваться в несколько ограниченном и идеализированном классе каналов. Ясно, что результаты, полученные для таких моделей, можно применять только к каналам, которые хорошо описываются этими моделями. Однако, изучая вероятность ошибки, мы интересуемся прежде всего весьма нетипичными событиями, вызывающими ошибки. Нелегко найти модель, хорошо описывающую как типичные, так и нетипичные события. Следовательно, анализ кодов в идеализированных каналах может дать только ограниченные сведения о поведении их в реальных каналах, и к результатам такого анализа следует подходить весьма осторожно.

Рассматриваемые здесь модели каналов называют симметричными каналами с двоичным входом. Мы имеем в виду канал с дискретным временем, входом которого служит последовательность двоичных символов

0 и 1, а выходом — соответствующая последовательность букв дискретного или непрерывного алфавита. Канал не обладает памятью, иначе говоря, при заданном в некоторый момент времени символе на входе соответствующий символ на выходе канала статистически не зависит от символов на входе и выходе канала в другие моменты времени. Требования симметрии будут точно определены в гл. 3, но, грубо говоря, они означают следующее: символы на выходе канала можно объединить в пары так, что вероятность одного выходного символа при заданном входном совпадает с вероятностью второго выходного символа при другом входном. Двоичный симметричный канал, сокращенно ДСК, принадлежит этому классу каналов; он имеет два символа на выходе, каждый из которых соответствует некоторому символу на входе. ДСК можно полностью охарактеризовать с помощью вероятности перехода одного входного символа в символ на выходе, соответствующий другому входному символу.

Если нужно использовать симметричный канал с двоичным входом без кодирования, передают последовательность двоичных символов и приемник восстанавливает символы на основе принятых, по одному в каждый момент времени. Если же для передачи по каналу используется кодирование, кодер должен сначала преобразовать последовательности двоичных символов, несущих информацию, в более длинные, содержащие избыточность последовательности, называемые кодовыми словами. Мы определим для таких кодов скорость R как отношение длин информационной последовательности и кодового слова. Если кодовые слова имеют длину n, то существует 2^{nR} возможных последовательностей на выходе источника, которые необходимо преобразовать в кодовые слова. Тогда в качестве кодовых слов можно использовать только $2^{-n(1-R)}$-ю долю из 2^n различных последовательностей длины n.

На приемном конце декодер по хранящимся в нем последовательностям, являющимся кодовыми словами, может отделить переданную последовательность

длины n от шума в канале. Так, кодовое слово отображается обратно на совокупность nR информационных символов. Во многих методах декодирования находят переданное слово, принимая сначала решение относительно каждого отдельного символа, а затем исправляют ошибки, пользуясь списком кодовых слов. Однако, как подробно показано для нескольких типов каналов в работе [1], при таком промежуточном решении теряется значительная доля информации о переданном сообщении. Описанный ниже

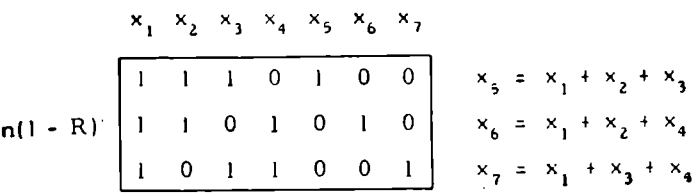


Рис. 1.2. Пример проверочной матрицы.

метод декодирования позволяет избежать промежуточного решения и оперирует непосредственно с *апостериорными* вероятностями входных символов, условными по отношению к соответствующим выходным символам.

Коды, рассматриваемые в данной работе, являются специальными случаями кодов с проверками на четность [1]). Кодовые слова в коде с проверками на четность образуются комбинированием блоков двоичных информационных символов с блоками проверочных символов. Каждый проверочный символ есть сумма по модулю 2 некоторой наперед указанной совокупности информационных символов [2]). Правила построения таких проверочных символов удобно представить в виде проверочной матрицы, такой, как приведенная на рис. 1.2. Такая матрица представляет собой систему линейных однородных уравнений по

[1]) Более подробное описание кодов с проверками на четность можно найти в работе Питерсона [12].

[2]) Сумма по модулю 2 равна 1, если обычная сумма нечетна, и равна 0, если обычная сумма четна.

модулю 2, называемых проверочными уравнениями, а множество кодовых слов есть множество решений этих уравнений. Совокупность символов, входящих в проверочное уравнение, мы назовем проверочным множеством. Например, первое проверочное множество на рис. 1.2 есть множество символов с индексами (1, 2, 3, 5).

Применение кодов с проверками на четность делает кодирование (в отличие от декодирования) сравнительно простым для реализации. Кроме того, как показал Элайес [3], если в ДСК используется типичный код с проверками на четность с большой длиной блока и если скорость кода лежит между критической скоростью и пропускной способностью канала, то вероятность ошибки декодирования почти совпадает с вероятностью ошибки для лучшего из возможных кодов с той же скоростью передачи и длиной блока.

К сожалению, для кодов с проверками на четность не проста реализация декодирования, поэтому мы вынуждены искать специальные классы таких кодов с проверками на четность, для которых существует приемлемый метод декодирования, например, класс кодов, описываемый в разд. 1.2.

1.2. Коды с малой плотностью проверок на четность

Коды с малой плотностью проверок на четность — это коды, определяемые матрицей, содержащей преимущественно нули и сравнительно небольшое число единиц. А именно (n, j, k)-код с малой плотностью проверок на четность есть код с длиной блока n и с матрицей, подобной приведенной на рис. 2.1, в которой каждый столбец содержит небольшое фиксированное число единиц j и каждая строка содержит небольшое фиксированное число единиц k. Заметим, что в матрицах этого типа проверочные символы расположены не по диагонали — в противоположность матрице, представленной на рис. 1.2; тем не менее для целей кодирования уравнения, соответствующие этим матрицам, всегда можно решить так, чтобы провероч-

ные символы были явными суммами информационных символов. Коды с малой плотностью не оптимальны в несколько искусственном смысле минимальности вероятности ошибки при заданной длине блока; больше того, можно показать, что максимально достижимая скорость этих кодов ограничена величиной, меньшей пропускной способности. Однако простота декодирования более чем компенсирует эти недостатки.

1.3. Сводка результатов

В гл. 2 будет построен ансамбль (n, j, k)-кодов, мы используем его для анализа свойств расстояний этих кодов. Расстояние между двумя кодовыми словами есть просто число символов, в которых они различаются. Ясно, что совокупность расстояний между некоторым словом и всеми остальными кодовыми словами есть важный параметр кода. Можно показать [12], что в коде с проверками на четность все кодовые слова имеют одинаковые наборы расстояний до других кодовых слов. Поэтому свойства расстояния в ансамбле можно описать, используя типичные числа кодовых слов, находящихся на разных расстояниях от слова, состоящего целиком из нулей. Найдено, что для типичного (n, j, k)-кода с $j \geqslant 3$ минимальное расстояние линейно возрастает с длиной блока при постоянных j и k. На рис. 2.4 приведено отношение минимального расстояния к длине блока при нескольких значениях j и k; оно сравнивается с соответствующим отношением для обычных кодов с проверками на четность; (n, j, k)-код при $j=2$ ведет себя совсем по-другому; показано, что минимальное расстояние $(n, 2, k)$-кода может расти самое большее логарифмически с длиной блока.

В гл. 3 найдена некоторая общая верхняя граница для вероятности ошибки декодирования в симметричном канале с двоичным входом при декодировании по максимуму правдоподобия как для кодов, так и для произвольных ансамблей кодов. Граница связана с кодом только через свойства расстояний. Допущение о декодировании по максимуму правдоподобия

введено отчасти из-за вносимого им упрощения вычислений, частично же для того, чтобы было можно оценивать коды независимо от алгоритма декодирования. Любой практически реализуемый алгоритм декодирования, такой, как описанный в гл. 4, приводит к необходимости выбора между малостью вероятности ошибки и простотой; декодирование по максимуму правдоподобия минимизирует вероятность ошибки, но абсолютно неприменимо, если длина блока велика.

В гл. 3 показано, что если расстояния кодовых слов связаны линейно с длиной блока и если скорость кода достаточно мала, граница $P(e)$ оказывается экспоненциально убывающей функцией длины блока. Для выбранных подходящим образом ансамблей кодов эти оценки сводятся к обычным оценкам случайного кодирования [3, 4].

Особенно простая оценка найдена, в частности, для двоичного симметричного канала. С ее помощью показано, что при всех вероятностях перехода в канале для типичного кода с малой плотностью поведение вероятностей ошибки такое же, как и для оптимального кода с немного большей скоростью. Рис. 3.5 иллюстрирует проигрыш в скорости, связанный с использованием кодов с малой плотностью проверок на четность.

В гл. 4 описаны два метода декодирования. В соответствии с первым, особенно простым методом декодер вначале принимает решение о каждом символе, а затем вычисляет проверки на четность и изменяет на обратные все символы, содержащиеся больше чем в некотором фиксированном числе неудовлетворившихся проверочных соотношений. Процесс повторяется до тех пор, пока последовательность не будет декодирована, причем каждый раз используются измененные символы. Второй метод декодирования основан на вычислении условных вероятностей того, что символ на входе равен 1. Эти вероятности вычисляются при условии, что известны все принятые символы, входящие в любое проверочное уравнение, содержащее рассматриваемый символ. И опять процесс повторяется до тех пор, пока последовательность не

будет декодирована. Число операций на символ при каждом повторении в обоих методах не зависит от длины кода. Второй, вероятностный метод требует несколько большего числа операций, однако позволяет декодировать с меньшей вероятностью ошибки.

Математический анализ вероятности появления ошибки при вероятностном методе декодирования труден из-за статистических зависимостей. Однако для ДСК с достаточно малой вероятностью перехода и для кодов с $j \geqslant 4$ удалось получить очень слабую оценку сверху вероятности ошибки; она убывает экспоненциально с убыванием корня из длины блока. На рис. 3.5 отложены вероятности перехода, для которых вероятность ошибки декодирования во всяком случае стремится к 0 с возрастанием длины кода. Высказывается гипотеза о том, что в действительности вероятность ошибки декодирования убывает экспоненциально с возрастанием длины блока, а число итераций, необходимых для декодирования, растет логарифмически.

Все основные результаты гл. 2, 3 и 4 распространяются в гл. 5 на недвоичные коды с малой плотностью проверок на четность. Хотя такое обобщение вполне естественно, выражения для минимального расстояния, вероятности ошибки и вероятности ошибки при вероятностном декодировании очень сложны, и поэтому очень мало можно сказать о преимуществах и недостатках недвоичных кодов по сравнению с двоичными. По-видимому, для оценки качества этих кодов окажется полезной дальнейшая экспериментальная работа.

Некоторые экспериментальные результаты о двоичных кодах с малой плотностью приведены в гл. 6. Для моделирования вероятностного декодирования и шумов, возникающих в каналах нескольких разных типов, использовалась вычислительная машина ИБМ-7090. Ввиду ограниченности машинного времени исследовался только случай каналов со значительным уровнем шумов, таких, что вероятность ошибки декодирования превышала 10^{-4}. На рис. 6.8 приведены наиболее поучительные результаты экспериментов,

они особенно подчеркивают преимущества метода декодирования с использованием приемника, вычисляющего отношение правдоподобия, по сравнению с решающим приемником.

1.4. Сравнение с другими методами

Перечислим другие методы кодирования и декодирования, которые представляются весьма перспективными с точки зрения получения малой вероятности ошибки и большой скорости передачи информации без особо больших затрат: первый — сверточные коды [3] с последовательным декодированием, развитым Возенкрафтом [17], Фано [5] и Рейффеном [14], второй — сверточные коды с пороговым декодированием Месси [10] и третий — коды Боуза — Чоудхури с декодированием по методу Питерсона [12] и Цирлера и Горенстейна [18].

Фано показал [5], что последовательное декодирование позволяет в любом дискретном канале без памяти декодировать с вероятностью ошибки, ограниченной сверху функцией вида $e^{-\alpha n}$. Здесь n — длина кодовых ограничений, а α — функция как канала, так и кода, причем α положительна для скоростей, меньших пропускной способности C. Фано показал также, что при скоростях, меньших некоторой величины R_{comp}, где $R_{\text{comp}}<C$, среднее число операций при декодировании символа ограничено величиной, не зависящей от длины кодовых ограничений.

В Линкольновской лаборатории (Лексингтон, Массачусетс) [11] построен экспериментальный декодер. Применение его в системе с обратной связью и с соответствующими модулятором и демодулятором позволило получить экспериментально надежную связь по телефонной линии со скоростью около 7500 *бит/сек*, в то время как без кодирования возможно достичь скорости всего в 1200 или 2400 *бит/сек*.

Последовательному декодированию присущи два основных недостатка: первый — число операций на символ есть случайная величина, это создает проблему очереди в декодере; второй — если декодер со-

вершает ошибку, возможно появление большого числа ошибок, прежде чем декодер вернется на правильный путь. Эти проблемы не вызывают серьезных затруднений при наличии обратной связи, однако случай, когда линия обратной связи отсутствует, требует дальнейшего исследования.

Пороговое декодирование — наиболее просто реализуемый из рассматриваемых здесь методов — требует наличия только регистров сдвига, некоторого числа двоичных сумматоров и порогового устройства. Оно наиболее эффективно при сравнительно небольших длинах кодовых ограничений, однако дает несколько бо́льшую вероятность ошибки и менее гибко, чем метод последовательного декодирования.

Число операций на символ при декодировании кодов Боуза — Чоудхури в двоичных симметричных каналах растет, грубо говоря, как куб длины блока, и слабо флуктуирует. Метод декодирования гарантирует исправление всех комбинаций с числом ошибок вплоть до некоторой величины и не позволяет исправлять никаких других комбинаций. Уже при блоках умеренной длины это ограничение вызывает большой рост P_e. Неизвестно, каким образом можно использовать апостериорные вероятности на выходе более общих каналов с двоичным входом. По-видимому, то, что алгебраические методы декодирования не позволяют использовать апостериорные вероятности, есть их характерное отличие от вероятностных методов.

Число операций на символ при декодировании кодов с малой плотностью проверок на четность, по-видимому, растет самое большее логарифмически с длиной блока и слабо флуктуирует. Вероятность ошибки декодирования неизвестна, однако есть основание считать, что она убывает экспоненциально с ростом длины блока. Возможность декодировать параллельно все символы позволяет получить скорости передачи информации большие, чем при использовании других методов.

Для многих каналов с памятью можно запоминать апостериорные вероятности на выходе канала, и это практически исключает необходимость принимать во

внимание память канала каким-либо иным способом. В канале с замираниями, например когда замирания охватывают несколько бодов подряд, апостериорные вероятности укажут на присутствие замирания. Если же этот канал используется как двоичный симметричный, при декодировании нужно учитывать тот факт, что пакеты ошибок более вероятны, чем изолированные ошибки. Таким образом, использование апостериорных вероятностей приводит к тому, что декодирование кодов с малой плотностью и последовательное декодирование оказываются более гибкими при работе в каналах с зависимыми шумами. С другой стороны, существует особенно простой метод декодирования кодов Боуза — Чоудхури [12] для того случая, когда действие шумов ограничивается появлением коротких, но плотных пакетов ошибок.

Когда передача по каналу происходит с длительными замираниями или подвергается действию длинных пакетов ошибок, часто практически не имеет смысла исправлять ошибки в эти периоды. В таких случаях использование сочетания исправления и обнаружения ошибок при обратной связи и переспросе имеет определенные преимущества. Все рассматриваемые здесь методы кодирования и декодирования естественным образом укладываются в эту схему, однако в тех случаях, когда исправление не осуществляется или исправляется небольшое число ошибок, коды с малой плотностью оказываются малоэффективными.

В заключение скажем, что все перечисленные методы имеют свои достоинства и недостатки и ясно, что ни один из них не может оказаться оптимальным во всех случаях. Сейчас, по-видимому, существует достаточное число вариантов методов кодирования и декодирования, чтобы можно было всерьез рассматривать вопрос об их использовании в конкретных каналах.

Глава 2

ФУНКЦИИ РАССТОЯНИЯ

Функция расстояния $N(l)$ для кода с проверками на четность определяется как число кодовых слов веса l. Из групповых свойств кода с проверками на четность легко следует [12], что $N(l)$ есть также число кодовых слов, находящихся на расстоянии l от любого кодового слова. Тогда определим минимальное расстояние кода D как наименьшее $l>0$, для которого $N(l)\neq 0$. Ясно, что в коде с заданной длиной блока n и со скоростью R желательно сделать D возможно бóльшим, а для тех l, которые превышают D, сделать $N(l)$ возможно меньшим. В следующей главе, где обсуждаются оценки вероятности ошибки декодирования в симметричных каналах с двоичным входом, станет более ясной точная связь $N(l)$ со способностью кода исправлять ошибки.

Для кодов большой длины с проверками на четность из-за огромного числа кодовых слов обычно практически невозможно точно вычислить функцию расстояния или даже минимальное расстояние. Часто проще исследовать среднюю по ансамблю кодов функцию расстояния, поскольку статистические свойства ансамбля допускают осреднение тех величин, которые не поддаются анализу в конкретном коде. По средним можно сделать некоторые выводы статистического характера об отдельных кодах, входящих в ансамбль.

2.1. Ансамбль равновероятных кодов с проверками на четность

В этой главе рассматриваются в основном функции расстояния кодов с малой плотностью проверок на четность, однако для сравнения мы получим

сначала осредненную функцию расстояния для другого ансамбля кодов с проверками на четность. Поскольку код с проверками на четность полностью задается проверочной матрицей, ансамбль кодов с проверками на четность можно определить посредством ансамбля проверочных матриц.

Ансамбль равновероятных кодов с проверками на четность со скоростью R и длиной блока n определим как ансамбль $n(1-R)\times n$ проверочных матриц, элементами которых являются независимые и равновероятные двоичные символы. По существу, этот ансамбль совпадает с ансамблем, рассматривавшимся Элайесом [3] при построении оценок случайного кодирования для кодов с проверками на четность. Незначительное отличие нашего ансамбля заключается в том, что входящие в него коды могут иметь скорость, немного превышающую R, поскольку строки матриц в этом ансамбле не обязательно линейно независимы над полем чисел по модулю 2.

Теорема 2.1. *Пусть $\overline{N(l)}$ есть среднее число кодовых слов веса l в ансамбле равновероятных кодов с проверками на четность со скоростью R и длиной блока n. Тогда для $l>0$ имеем*

$$\overline{N(l)} = C_n^l 2^{-2(1-R)} \leqslant$$

$$\leqslant [2\pi n\lambda(1-\lambda)]^{-1/2} \exp n[H(\lambda) - (1-R)\ln 2], \qquad (2.1)$$

где

$$\lambda = \frac{l}{n},$$

$$H(\lambda) = \lambda \ln \frac{1}{\lambda} + (1-\lambda)\ln \frac{1}{1-\lambda}.$$

Доказательство. Пусть $P(l)$ есть вероятность множества кодов, для которых некоторая фиксированная последовательность веса l есть кодовое слово. Иначе говоря, $P(l)$ есть вероятность того, что фиксированная последовательность веса l окажется кодовым словом кода, выбранного случайным образом из ансамбля. Поскольку последовательность, це-

ликом состоящая из нулей, принадлежит любому коду с проверками на четность, то $P(l)=1$ при $l=0$. При $l\neq 0$ проверочное соотношение с вероятностью $1/2$ включает позицию, в которой находится последняя из единиц последовательности веса l. Поэтому независимо от того, входили первые $l-1$ единиц в проверочное соотношение четное или нечетное число раз, вероятность того, что соотношение будет удовлетворено, равна $1/2$. Последовательность будет кодовым словом тогда и только тогда, когда она удовлетворяет всем $n(1-R)$ проверочным соотношениям; таким образом,

$$P(l)=2^{-n(1-R)} \quad \text{при} \quad l\neq 0.$$

Вероятность $P(l)$ можно также интерпретировать как математическое ожидание случайной величины, принимающей значение 1, если последовательность принадлежит коду, и 0 в противном случае. Заметим теперь, что всего существует C_n^l последовательностей веса l и что среднее число кодовых слов среди этих последовательностей равно сумме вероятностей того, что каждая последовательность принадлежит коду; таким образом,

$$\overline{N(l)}=C_n^l 2^{-n(1-R)}. \tag{2.2}$$

Оценим теперь C_n^l по формуле Стирлинга

$$\frac{1}{\sqrt{2\pi n}}n^n\exp\left(-n+\frac{1}{12n}-\frac{1}{360n^3}\right)\leqslant$$
$$\leqslant n!\leqslant\frac{1}{\sqrt{2\pi n}}n^n\exp\left(-n+\frac{1}{12n}\right). \tag{2.3}$$

После некоторых преобразований получим для $\lambda n=l$

$$\frac{1}{\sqrt{2\pi n\lambda(1-\lambda)}}\exp\left[nH(\lambda)-\frac{1}{12n\lambda(1-\lambda)}\right]<$$
$$<C_n^{n\lambda}<\frac{1}{\sqrt{2\pi n\lambda(1-\lambda)}}\exp nH(\lambda), \tag{2.4}$$

где

$$H(\lambda)=-\lambda\ln\lambda-(1-\lambda)\ln(1-\lambda).$$

Объединяя соотношения (2.2) и (2.4), получим утверждение теоремы. Ч.Т.Д.

Заметим теперь, что в ансамбле равновероятных кодов с проверкой на четность минимальное расстояние кода есть случайная величина; ее функция распределения оценивается в следующей теореме.

Теорема 2.2. *В ансамбле равновероятных кодов с проверками на четность со скоростью R и длиной блока n функция распределения минимального расстояния* $\Pr(D \leqslant \delta n)$ *оценивается при* $\delta < 1/2$ *и целом* δn *двумя следующими неравенствами:*

$$\left.\begin{aligned} &\Pr(D \leqslant \delta n) \leqslant \\ &\leqslant \frac{1}{1-2\delta}\sqrt{\frac{1-\delta}{2\pi n\delta}}\exp n[H(\delta) - (1-R)\ln 2], \\ &\Pr(D \leqslant \delta n) \leqslant 1. \end{aligned}\right\} \quad (2.5)$$

Доказательство. В теореме 2.1 было показано, что в ансамбле кодов вероятность того, что ненулевая последовательность есть кодовое слово, равна $2^{-n(1-R)}$. Вероятность того, что некоторая последовательность веса $n\delta$, или меньше, есть кодовое слово, безусловно, меньше суммы вероятностей того, что каждая из последовательностей является кодовым словом. Поэтому

$$\Pr(D \leqslant n\delta) \leqslant \sum_{l=1}^{n\delta} C_n^l 2^{-n(1-R)} \times$$

$$\times \sum_{l=1}^{n\delta} C_n^l = C_n^{n\delta}\left[1 + \frac{n\delta}{n-n\delta+1} + \right.$$

$$\left. + \frac{n\delta(n\delta-1)}{(n-n\delta+1)(n-n\delta+2)} + \ldots\right]. \quad (2.6)$$

Это выражение мажорируется геометрической прогрессией, и мы получаем таким образом

$$\sum_{l=1}^{n\delta} C_n^l \leqslant C_n^{n\delta}\left(\frac{1-\delta}{1-2\delta}\right). \quad (2.7)$$

Оценив правую часть неравенства (2.7) с помощью неравенства (2.4) и подставив результат в неравенство (2.6), получим утверждение теоремы. Ч.Т.Д.

С ростом n эта оценка $\Pr(D \leqslant n\delta)$ как функция δ сходится к ступенчатой функции со скачком при $\delta_0 < 1/2$, для которого $H(\delta_0) = (1-R)\ln 2$. Зависимость δ_0 от скорости приведена на рис. 2.4. Этот результат тесно связан с границей Гилберта для минимального расстояния [1]) [6].

Асимптотическая форма границы Гилберта утверждает, что для больших n существует код, для которого $D \geqslant n\delta_0$. Теорема 2.2 утверждает, что для любого $\varepsilon > 0$ вероятность множества кодов с проверками на четность, для которых $D < n(\delta_0 - \varepsilon)$, стремится к 0 экспоненциально с ростом n.

2.2. Свойства расстояния кодов с малой плотностью проверок

В этом разделе мы определим ансамбль кодов с малой плотностью проверок на четность и докажем теоремы, аналогичные теоремам 2.1 и 2.2. Затем мы построим новый ансамбль, отбрасывая коды с малым минимальным расстоянием. Такой улучшенный ансамбль будет использован в следующей главе при выводе оценок вероятности ошибки декодирования для различных каналов.

Определим (n, j, k)-проверочную матрицу как матрицу с n столбцами, j единицами в каждом столбце, k единицами в каждой строке и с нулями во всех остальных позициях. Из определения следует, что (n, j, k)-проверочная матрица имеет nj/k строк и поэтому скорость $R \geqslant 1 - j/k$. Для того чтобы построить ансамбль (n, j, k)-матриц, рассмотрим сначала (n, j, k)-матрицу, приведенную на рис. 2.1, для которой $n = 20$, $j = 3$ и $k = 4$.

---

[1]) Границу Гилберта принято называть границей Варшамова — Гилберта. См., например, [12]. — *Прим. перев.*

Эта матрица разбита на j подматриц, и в каждом столбце каждой подматрицы содержится только одна единица. В первой подматрице все единицы расположены в ступенчатом порядке, т. е. i-я строка содержит единицы в столбцах с $[(i-1)k+1]$-го по ik-й. Остальные подматрицы суть просто перестановки

1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0
0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0
0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0
0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1

Рис. 2.1. Пример матрицы кода с малой плотностью проверок на четность; $n=20$, $k=4$, $j=3$.

столбцов первой. Определим ансамбль (n, j, k)-кодов как ансамбль, получающийся при случайных перестановках столбцов каждой из $(j-1)$ нижних подматриц матрицы типа приведенной на рис. 2.1, причем всем перестановкам приписываются равные вероятности. Это определение несколько произвольно и введено для удобства математического анализа. На самом деле такой ансамбль не включает все только что определенные (n, j, k)-коды. К тому же по крайней мере $(j-1)$ строк каждой матрицы ансамбля

линейно зависимы. Это означает просто, что скорость кода немного больше той, которая определяется матрицей.

Прежде чем искать среднюю функцию расстояния и функцию распределения минимального расстояния для этих ансамблей кодов, докажем следующую необходимую нам теорему.

Теорема 2.3. *Для каждого кода из* (n, j, k)*-ансамбля число* $N_1(l)$ *последовательностей веса* l, *удов-*

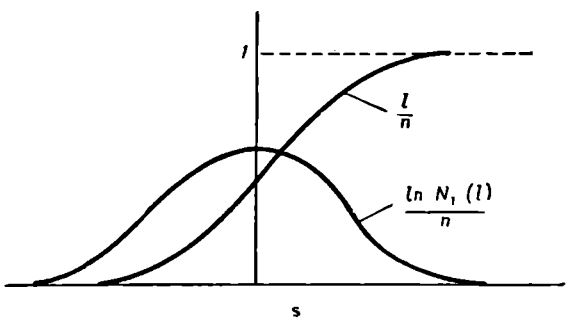


Рис. 2.2. Зависимость функций $\frac{l}{n}$ и $\frac{\ln N_1(l)}{n}$ от параметра S.

летворяющих какому-либо из j *блоков* n/k *проверочных уравнений, оценивается следующим образом*:

$$N_1\left[\frac{n}{k}\mu'(s)\right] \leqslant \exp\frac{n}{k}[\mu(s) - s\mu'(s) + (k-1)\ln 2], \quad (2.8)$$

где s — *произвольный параметр,* $\mu(s)$ *определяется как*

$$\mu(s) = \ln 2^{-k}[(1+e^s)^k + (1-e^s)^k] \quad (2.9)$$

и

$$\mu'(s) = \frac{d\mu(s)}{ds}.$$

Эта теорема связывает l и $N_1(l)$, выражая их как функции параметра s. На рис. 2.2 приведена зависимость l/n и $[\ln N_1(l)]/n$ от s.

Доказательство. Для любого кода в ансамбле и любого из j блоков по n/k проверочных соотношений n/k проверочных множеств [1]) не пересекаются и исчерпывают все позиции блока. Рассмотрим совокупность всех двоичных последовательностей длины k, содержащих четное число единиц, и построим из них ансамбль, приписывая всем им равные вероятности. Общее число последовательностей в ансамбле равно 2^{k-1}, а вероятность того, что последовательность содержит i единиц (i четно), равна $C_k^i 2^{-k+1}$. Отсюда выражение для производящей функции моментов числа единиц в последовательности будет

$$g(s)=\sum_{\substack{\text{четные}\\ i}} C_k^i 2^{-k+1} e^{si}. \qquad (2.10)$$

или

$$g(s)=2^{-k}[(1+e^s)^k+(1-e^s)^k]. \qquad (2.11)$$

Для того чтобы показать, что правые части равенств (2.10) и (2.11) эквивалентны, нужно разложить правую часть равенства (2.11) по формуле бинома и заметить, что нечетные члены уничтожаются.

Для каждого из n/k проверочных множеств выберем независимо последовательность из построенного выше ансамбля и используем ее в качестве двоичных единиц в позициях этого проверочного множителя. Таким образом определим ансамбль равновероятных событий, в котором каждое событие есть последовательность длины n, удовлетворяющая n/k проверочным соотношениям. Число единиц в каждой последовательности длины n есть сумма чисел единиц, входящих в каждое проверочное множество, и, таким образом, есть сумма n/k независимых случайных величин, причем производящая функция моментов каждой из них определяется равенством (2.11). Следовательно, производящая функция моментов числа единиц в последовательности длины n есть $[g(s)]^{n/k}$. Используем теперь это обстоятельство для оценки ве-

[1]) То есть множеств позиций, для которых коэффициенты в проверочном соотношении отличны от нуля. — *Прим. ред.*

роятности $Q(l)$ того, что последовательность из ансамбля содержит l единиц. По определению,

$$[g(s)]^{\frac{n}{k}} = \sum_{l=0}^{n} \exp(sl)\, Q(l) \geqslant \tag{2.12}$$

$$\geqslant \exp(sl)\, Q(l) \tag{2.13}$$

для любых s и l.

Из равенств (2.9) и (2.11) следует, что $\mu(s) = \ln g(s)$ и

$$Q(l) \leqslant \exp\left[\frac{n}{k}\mu(s) - sl\right]:$$

И наконец, $N_1(l)$ равно произведению $Q(l)$ на число последовательностей в ансамбле. Поскольку в ансамбле последовательностей длины k всего 2^{k-1} элементов, имеется $2^{n(k-1)/k}$ элементов в ансамбле последовательностей длины n, отсюда

$$N_1(l) \leqslant \exp\left[\frac{n}{k}\mu(s) + \frac{n}{k}(k-1)\ln 2 - sl\right]. \tag{2.14}$$

Положим производную экспоненты в неравенстве (2.14) равной 0 и подставим полученное значение $l = (n/k)\mu'(s)$ в неравенство (2.14). Это даст нам неравенство (2.8), доказывающее теорему. Ч.Т.Д.

Как показано в [4], положив $l = (n/k)\mu'(s)$, мы действительно минимизируем экспоненту, получая таким образом оптимальную оценку; теорема, однако, верна безотносительно к минимальности экспоненты. Можно показать (хотя в этом нет необходимости при нашем доказательстве), используя «перекошенные» распределения [4] и центральную предельную теорему [7], что при больших n справедливо асимптотическое равенство

$$N_1\left[\frac{n}{k}\mu'(s)\right] \to \frac{2}{\sqrt{2\pi n\mu'(s)}} \times$$

$$\times \exp\frac{n}{k}[\mu(s) - s\mu'(s) - (k-1)\ln 2]. \tag{2.15}$$

Мы можем теперь воспользоваться теоремой 2.3 для того, чтобы найти вероятность $P(l)$ множества

кодов, для которых некоторая фиксированная последовательность веса l есть кодовое слово. Ясно, что, поскольку все перестановки кода равновероятны, $P(l)$ не зависит от конкретного вида такой последовательности.

Если последовательность веса l выбрана случайным образом, то для любого кода из ансамбля вероятность того, что эта последовательность удовлетворит любому фиксированному блоку из n/k проверочных соотношений, равна $N_1(l)/C_n^l$. Поскольку все j блоков проверочных соотношений выбраны независимо, то

$$P(l)=\left[\frac{N_1(l)}{C_n^l}\right]^j. \tag{2.16}$$

Теперь мы можем выразить через $P(l)$ свойства расстояния и функцию распределения минимального расстояния, так же как это было сделано для ансамбля всех кодов с проверками на четность в соотношениях (2.1) и (2.5):

$$\overline{N_{jk}(l)} \leqslant C_n^l P(l)=(C_n^l)^{-j+1}[N_1(l)]^j. \tag{2.17}$$

$$\Pr(D \leqslant n\delta) \leqslant \sum_{l=2}^{n\delta} C_n^l P(l)=\sum_{l=2}^{n\delta}(C_n^l)^{j+1}[N_1(l)]^j. \tag{2.18}$$

Заметим, что в ансамбле кодов с малой плотностью кодовыми словами могут быть только последовательности четного веса. Используя неравенства (2.4) и (2.14), получаем

$$\overline{N_{jk}(t)} \leqslant C(\lambda,\ n)\exp -nB_{jk}(\lambda),$$

где

$$\lambda=\frac{l}{n}; \tag{2.19}$$

$$B_{jk}(\lambda)=(j-1)H(\lambda)-\frac{j}{k}[\mu(s)+(k-1)\ln 2]+js\lambda; \tag{2.20}$$

$$C(\lambda,\ n)=[2\pi n\lambda(1-\lambda)]^{\frac{j-1}{2}}\exp\frac{j-1}{12n\lambda(1-\lambda)}, \tag{2.21}$$

где $\lambda=\mu'(s)/k$.

Подставляя неравенство (2.19) в неравенство (2.18), получаем

$$\Pr(D \leqslant n\delta) \leqslant \sum_{l=2}^{n\delta} C(\lambda, n) \exp[-nB_{jk}(\lambda)]. \quad (2.22)$$

При больших n суммы в неравенствах (2.19) и (2.22) определяются главным образом поведением $B_{jk}(\lambda)$; функция $B_{jk}(\lambda)$ появляется также в следующей главе в оценках вероятности ошибки декодирования. К сожалению, исследовать $B_{jk}(\lambda)$ нелегко, поскольку эта

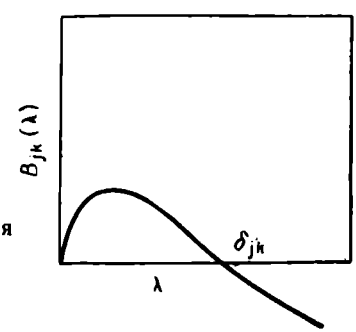


Рис. 2.3. Пример поведения функции $B_{jk}(\lambda)$.

величина выражается через параметр s, в свою очередь неявно зависящий от λ. В приложении А показано, что $B_{jk}(\lambda)$ для $j \geqslant 3$ ведет себя так, как показано на рис. 2.3. Она равна 0 при $\lambda = 0$, затем растет, причем производная в нуле бесконечна, достигает максимума, затем убывает, пересекает ось абсцисс при некотором $\lambda = \delta_{jk}$ и остается отрицательной при $\lambda > \delta_{jk}$.

Ясно, что для любого $\delta > \delta_{jk}$ сумма в неравенстве (2.22) не ограничена, однако функцию распределения минимального расстояния все же можно оценить сверху единицей. При $\delta < \delta_{jk}$ наибольшие слагаемые имеют индексы λ, близкие к 0 и к δ_{jk}. Это четко сформулировано в следующей теореме, доказательство которой приводится в приложении А.

Теорема 2.4. *В* (n, j, k)*-ансамбле кодов функция распределения минимального расстояния оценивается следующими двумя неравенствами:*

$$\Pr(D \leqslant n\delta) \leqslant \frac{k-1}{2n^{j-2}} + o\left(\frac{1}{n^{j-2}}\right) + \\ + nC(\delta n, n)\exp[-nB_{jk}(\delta)] \qquad (2.23)$$

и

$$\Pr(D \leqslant n\delta) \leqslant 1.$$

где C *и* B *определяются равенствами* (2.20) *и* (2.21), *а* $o(1/n^{j-2})$ *— функция, убывающая с ростом* n *быстрее, чем* $1/n^{j-2}$.

Первое слагаемое в неравенстве (2.23) соответствует кодовым словам веса 2, следующее слагаемое соответствует словам с малыми весами, большими 2, а последнее соответствует словам с большим весом. При возрастании $\delta = 2/n$ такая оценка функции распределения минимального расстояния сходится к ступенчатой функции с небольшим скачком при $\delta = 2/n$ и большим скачком при $\delta = \delta_{jk}$, причем величина меньшего скачка убывает как n^{-j+2}.

Будем называть выражение δ_{jk} *типичным коэффициентом минимального расстояния* в (n, j, k)-ансамбле. При больших n подавляющая часть кодов в ансамбле имеет минимальное расстояние, либо близкое к $n\delta_{jk}$, либо большее; поскольку δ_{jk} не зависит от длины блока, минимальное расстояние, характерное для большинства кодов ансамбля, растет линейно с ростом длины блока. На рис. 2.4 приведена зависимость δ_{jk} от скорости для некоторых значений j и k, и эта величина сравнивается с коэффициентом минимального расстояния для ансамбля равновероятных кодов. Можно заметить, что с ростом j и k для (n, j, k)-кодов δ_{jk} быстро приближается к δ_0 ансамбля равновероятных кодов. Это доказывается в теореме А.3 приложения А.

Здесь же мы видим, зачем функция распределения минимального расстояния была получена раньше каких-либо результатов о вероятности ошибки декодирования. Если два слова в групповом коде отличают-

ся только в двух символах, вероятность ошибки декодирования оценивается снизу вероятностью принять эти два символа неправильно, а эта вероятность не зависит от длины кода. Таким образом, осредненная по ансамблю вероятность ошибки декодирования при $n \to \infty$ пропорциональна $1/n^{j-2}$, иначе говоря, пропорциональна вероятности кода с минимальным расстоя-

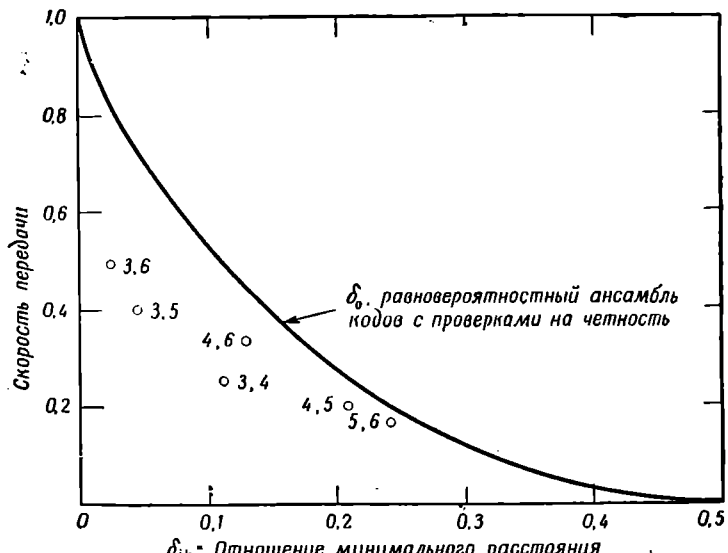


Рис. 2.4. Отношение минимального расстояния к длине блока у типичного (n, j, k)-кода большой длины.

нием 2. Итак, очень небольшое число плохих кодов определяет в основном вероятность ошибки, осредненную по ансамблю.

Теперь модифицируем (n, j, k)-ансамбль для того, чтобы определить вероятность ошибки типичного (n, j, k)-кода с минимальным расстоянием порядка $n\delta_{jk}$. Удалим из (n, j, k)-ансамбля половину кодов с наименьшим минимальным расстоянием и удвоим вероятность каждого из оставшихся кодов. Полученный ансамбль будем называть *улучшенным ансамблем*;

он используется в гл. 3 при выводе оценок вероятности ошибки декодирования (n, j, k)-кода.

Пусть δ_{njk} есть минимальное расстояние в улучшенном ансамбле. Тогда δ_{njk} оценивается снизу тем значением δ, при котором правая часть неравенства (2.23) равна половине. С ростом n оценка, даваемая неравенством (2.23), сходится к функции со скачком при δ_{jk}; таким образом, δ_{njk} асимптотически оценивается с помощью δ_{jk}. Мы имеем поэтому для улучшенного ансамбля кодов с малой плотностью следующую оценку:

$$\overline{N_{jk}(l)}\begin{cases}\leqslant 2C(\lambda,\ n)\exp\left[-nB_{jk}(\lambda)\right]; & \lambda\geqslant\delta_{njk},\\ =0; & \lambda<\delta_{njk}.\end{cases} \tag{2.24}$$

Проведя аналогичным образом улучшение ансамбля случайных кодов с проверками на четность, получаем из соотношений (2.1) и (2.5)

$$\overline{N(l)}\begin{cases}\leqslant 2\left[2\pi n\lambda(1-\lambda)\right]^{-1/2}\times\\ \qquad\times\exp n\left[H(\lambda)-(1-R)\ln 2\right]; & \lambda>\delta_0,\\ =0; & \lambda\leqslant\delta_0,\end{cases} \tag{2.25}$$

где δ_0 удовлетворяет уравнению $H(\delta_0)=(1-R)\ln 2$, а n достаточно велико, так что выполняется неравенство

$$\frac{1}{1-2\delta_0}\sqrt{\frac{1-\delta_0}{2\pi n\delta_0}}\leqslant\frac{1}{2}.$$

Прежде чем использовать этот модифицированный ансамбль для вывода оценок вероятности ошибки декодирования, рассмотрим частный случай $j=2$, соответствующий ансамблям, в которых каждый символ входит в точности в два проверочных множества.

Теорема 2.5. *Пусть в коде с проверками на четность длина блока равна n, каждый символ входит в точности в два проверочных множества и каждое проверочное множество содержит k символов. Тогда*

минимальное расстояние кода D оценивается следующим образом:

$$D \leqslant 2 + \frac{2 \ln \frac{n}{2}}{\ln (k-1)}. \qquad (2.26)$$

Доказательство. Докажем теорему, представив код в виде дерева — так, как это сделано на рис. 2.5. Пусть первый символ в коде представлен

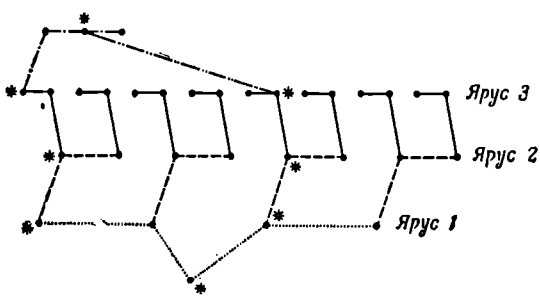


Рис. 2.5. Проверочное дерево.

корнем дерева. Этот символ входит в два проверочных множества, обозначенных двумя ветвями, выходящими из корня. Остальные символы этих двух проверочных множеств представлены узлами первого яруса дерева. Точно так же каждый символ первого яруса содержится в некотором другом проверочном множестве, представляемом ветвью, выходящей из этого символа.

Подобным образом можно строить ярусы дерева до тех пор, пока при некотором целом m ветвями, выходящими из m-го яруса, не будет образована петля. Такая петля может появиться либо тогда, когда два проверочных множества, выходящих из m-го яруса, содержат общий символ в $(m+1)$-м ярусе, как это показано на рис. 2.5, либо когда одно проверочное множество, выходящее из m-го яруса, содержит более одного символа в m-м ярусе.

Теперь оценим m в зависимости от длины блока n. Первый ярус дерева содержит $2(k-1)$ узлов,

второй — $2(k-1)^2$ узлов, аналогичным образом m-й ярус содержит $2(k-1)^m$ узлов, так как по предположению ветви ниже m-го яруса не образовывали петель. Поскольку каждому узлу соответствует свой символ,

$$2(k-1)^m \leqslant n,$$

$$m \leqslant \frac{\ln \frac{n}{2}}{\ln (k-1)}. \qquad (2.27)$$

Для фиксированной петли в дереве рассмотрим совокупность узлов на сочленениях ветвей петли. На рис. 2.5 эти узлы отмечены звездочками. Каждая ветвь петли должна содержать ровно два таких узла, и никакая из других ветвей дерева их не содержит. Поэтому последовательность длины n, содержащая единицы в позициях, соответствующих узлам нашего множества, и нули во всех других позициях, должна быть кодовым словом, так как все проверочные множества содержат четное число единиц. И наконец, вес кодового слова D, соответствующего первой из встретившихся петель, всегда оценивается следующим образом:

$$D \leqslant 2m+2, \qquad (2.28)$$

поскольку петля образована одним путем, идущим вверх по дереву, и одним путем, идущим вниз по дереву. Объединяя неравенства (2.27) и (2.28), получаем неравенство (2.26). Ч.Т.Д.

Глава 3

ВЕРОЯТНОСТЬ ОШИБКИ ДЕКОДИРОВАНИЯ

В этой главе развивается методика, позволяющая оценивать сверху вероятность ошибки декодирования произвольных двоичных блоковых кодов. Мы примем, что декодирование производится по методу максимума правдоподобия, а канал имеет двоичный алфавит на входе, произвольный алфавит на выходе и симметричен в некотором, определяемом ниже смысле.

Имеются три причины для построения этой методики. Во-первых, она позволяет выявить возможности кодов с малой плотностью проверок; во-вторых, она служит инструментом для сравнения кодов и лучшего понимания соотношения между свойствами кодового расстояния и вероятностью ошибки декодирования; в-третьих, она дает нам в руки идейно более простую, хотя и более сложную аналитически, технику анализа ансамблей случайных кодов. Идейная простота состоит в раздельном анализе канала и ансамбля кодов (используемом при получении свойств расстояния в ансамбле).

3.1. Симметричный канал с двоичным входом

Определим симметричный канал с двоичным входом как канал с дискретным временем, обладающий следующими свойствами:

1. Входной алфавит X состоит из двух символов, обозначаемых 0 и 1.
2. Выходной алфавит Y может представлять собой либо дискретное, либо непрерывное множество действительных чисел.

3. В каждый из дискретных моментов времени выход y статистически зависит только от входа x в тот же момент времени [1]).

4. Для выхода y выполнены условия симметрии, задаваемые следующим равенством:

$$P_0(y) = P_1(-y). \tag{3.1}$$

В этом равенстве и во всей главе $P_x(y)$ обозначает *условную плотность распределения*, если Y — непрерывное множество, и *условную вероятность*, если Y — дискретное множество.

На рис. 3.1 приведены некоторые примеры таких каналов. К сожалению, симметрия в обозначениях входов и выходов отсутствует. Изменение обозначений выходов сильно усложнило бы условие симметрии, задаваемое равенством (3.1), а изменение обозначений входов сделало бы коды с проверками на четность менее привычными для читателя, привыкшего к обозначению выходов символами 0 и 1.

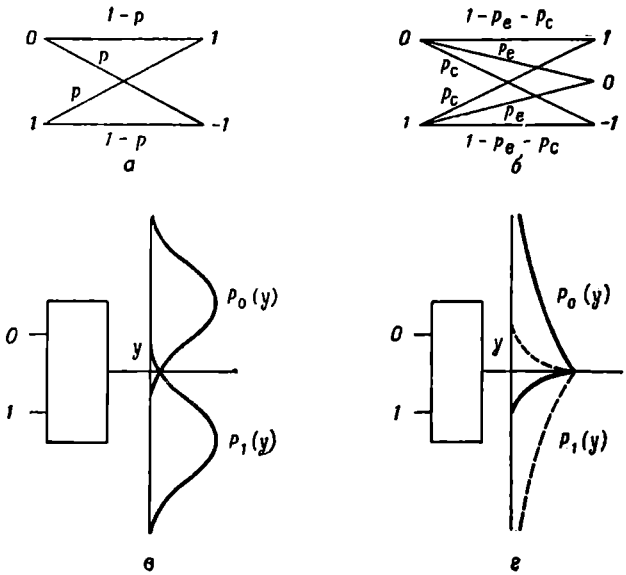
3.2. Свойства расстояния

Пусть передается произвольное слово u_0 некоторого кода с блоком длины n, и пусть известно число $N(l)$ других кодовых слов, находящихся на каждом расстоянии l от u_0. В следующем разделе оценивается сверху вероятность ошибки при декодировании по методу максимума правдоподобия, когда u_0 передается по симметричному каналу с двоичным входом. Эта оценка легко переносится на весь код или на ансамбль кодов.

3.3. Верхняя оценка вероятности ошибки декодирования

Пусть $u_0 = x_{10}, x_{20}, \ldots, x_{n0}$ — переданное кодовое слово, а $v = y_1, y_2, \ldots, y_n$ — принятая последовательность. Пусть $u_1, \ldots, u_j, \ldots, u_{M-1}$ — остальные кодо-

[1]) То есть при фиксированном входе во все моменты времени условное распределение y зависит лишь от x. — *Прим. ред.*



Р и с. 3.1. Симметричные каналы с двоичным входом.

а — двоичный симметричный канал;
б — двоичный симметричный канал с порогом;
в — канал с аддитивным гауссовским шумом, выход — логарифмическое правдоподобие (см. разд. 6.3)

$$P_0(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp - \frac{\left(y - \frac{\sigma^2}{2}\right)^2}{2\sigma^2};$$

$$P_1(y) = \frac{1}{\sqrt{2\pi\sigma}} \exp - \frac{\left(y + \frac{\sigma^2}{2}\right)^2}{2\sigma^2};$$

$$\sigma^2 = \frac{4E_c(1-\rho)}{N_0};$$

г — канал с релеевским замиранием, выход — логарифмическое правдоподобие (см. разд. 6.4)

$$P_0(y) = \frac{1+A}{A(2+A)} \exp - \frac{y}{A}; \quad y \geqslant 0;$$

$$P_0(y) = \frac{1+A}{A(2+A)} \exp - \frac{y(1+A)}{A}; \quad y < 0;$$

$$A = \frac{E_c}{N_0}.$$

вые слова, где $u_j = x_{1j}, x_{2j}, \ldots, x_{nj}$. Ошибка при декодировании по методу максимума правдоподобия происходит, если $P(v|u_j) > P(v|u_0)$ для некоторого j, $1 \leqslant j \leqslant M-1$. Ошибка декодирования может также произойти при $P(v|u_j) = P(v|u_0)$. Оценивая вероятность ошибки декодирования сверху, мы можем считать, что в таком случае всегда происходят ошибки. Основываясь на допущении о независимости n использований канала, можно записать условие для того, чтобы имела место ошибка

$$\prod_{l=1}^{n} P_{x_{lj}}(y_l) \geqslant \prod_{l=1}^{n} P_{x_{l0}}(y_l) \tag{3.2}$$

для некоторого j при $1 \leqslant j \leqslant M-1$.

Вероятность ошибки декодирования можно, таким образом, оценить сверху вероятностью выполнения неравенства (3.2). С неравенством (3.2) легче иметь дело, если мы возьмем логарифмы от обеих его частей; при этом получается следующее неравенство для сумм случайных величин:

$$\sum_{l=1}^{n} \ln P_{x_{lj}}(y_l) \geqslant \sum_{l=1}^{n} \ln P_{x_{l0}}(y_l). \tag{3.3}$$

Кроме того, по причинам, обсуждаемым ниже, мы вычтем из обеих частей неравенства (3.2) некоторую произвольную функцию выхода $\sum_{l=1}^{n} \ln f(y_l)$ и умножим обе части на -1, получая при этом следующее условие для того, чтобы имела место ошибка декодирования:

$$\sum_{l=1}^{n} \ln \frac{f(y_l)}{P_{x_{lj}}(y_l)} \leqslant \sum_{l=1}^{n} \ln \frac{f(y_l)}{P_{x_{l0}}(y_l)} \tag{3.4}$$

для некоторых j, $1 \leqslant j \leqslant M-1$. Наложим следующее ограничение на $f(y)$: $f(y)$ положительна, если положительна $P_0(y)$ или $P_1(y)$ и

$$f(y) = f(-y) \text{ для всех } y. \tag{3.5}$$

Определим теперь расстояние $\delta(x_i y_i)$ между входом x_i и выходом y_i:

$$\delta(x_i y_i) = \ln \frac{f(y_i)}{P_{x_i}(y_i)}. \tag{3.6}$$

Определим, кроме того, расстояние $D(uv)$ между u и v:

$$D(uv) = \sum_{i=1}^{n} \delta(x_i y_i). \tag{3.7}$$

Из соотношений (3.4), (3.6) и (3.7) нетрудно видеть, что ошибка декодирования происходит только тогда, когда $D(u_j v) \leqslant D(u_0 v)$ для некоторого u_j, отличного от u_0. Точнее, вероятность ошибки декодирования P_e оценивается следующим образом:

$$P_e \leqslant \Pr\left\{ \bigcup_{j=1}^{M-1} [\text{событие, заключающееся в том, что } D(u_j v) \leqslant D(u_0 v)] \right\}. \tag{3.8}$$

Наиболее очевидный метод упрощения неравенства (3.8) заключается в оценке сверху вероятности объединения событий суммой вероятностей событий. Это, однако, не дает хорошей оценки, поскольку в том случае, когда расстояние $D(u_0 v)$ очень велико, скажем больше некоторой соответствующим образом выбранной константы nd, весьма вероятно, что оно больше большинства величин $D(u_j v)$, поэтому ошибка декодирования учитывается в этой оценке много раз. Чтобы обойти эту трудность, мы будем отдельно оценивать события, для которых $D(u_0 v) \geqslant nd$. Параметр d произволен, и оптимизация по нему будет проведена позже. Разобьем неравенство (3.8) следующим образом:

$$P_e \leqslant P_1 + P_2, \tag{3.9}$$

где

$$P_1 = \Pr\left\{ \bigcup_{j=1}^{M-1} [\text{событие, заключающееся в том, что } D(u_0 v) > nd;\ D(u_j v) \leqslant D(u_0 v)] \right\};$$

$$P_2 = \Pr\left\{ \bigcup_{j=1}^{M-1} [\text{событие, заключающееся в том, что } D(u_0 v) \leqslant nd;\ D(u_j v) \leqslant D(u_0 v)] \right\}.$$

Теперь мы можем отдельно оценить P_1 и P_2:

$$P_1 \leqslant \Pr[D(u_0 v) > nd], \tag{3.10}$$

$$P_2 \leqslant \sum_{j=1}^{M-1} \Pr[D(u_0 v) \leqslant nd;\ D(u_j v) \leqslant D(u_0 v)]. \tag{3.11}$$

Заметим, что неравенство (3.9) есть точное выражение для P_e, если не считать того, что мы приняли, что неопределенность (т. е. случай, когда $D(u_0 v) = D(u_j v)$) всегда вызывает ошибку. Таким образом, произвольность выбора $f(y)$ не оказывает влияния на неравенство (3.9), поскольку от этого выбора не зависит, какое слово декодируется, когда передано слово u_0. Однако выбор $f(y)$ влияет на неравенства (3.10) и (3.11), поскольку эта функция определяет множество выходных последовательностей v, для которых $D(u_0 v) \geqslant nd$.

Заметим, наконец, что $D(u_0 v)$ и $D(u_j v)$ определены в равенстве (3.7) как суммы случайных величин, поэтому задача оценки P_e сведена с помощью неравенств (3.10) и (3.11) к задаче оценки хвостов распределений сумм случайных величин. Это лучше всего делается с помощью оценок по методу Чернова, краткое изложение которого можно найти в приложении В. Более подробно они даны в работе Фано [4] (гл. 8).

3.4. Оценки Чернова

Для оценки P_e в неравенстве (3.10) нам необходима теорема, доказанная в приложении Б.

Теорема 3.1. *Пусть* $Z = \sum_{i=1}^{n} z_i$ — *сумма* n *независимых случайных величин, пусть* $P_i(z_i)$ — *плотность распределения* i-*й случайной величины, и пусть*

$g_i(s) = \int_{-\infty}^{\infty} \exp(sz_i) P_i(z_i) dz_i$ *есть производящая функция моментов i-й случайной величины. Тогда*

$$\Pr(Z \geqslant nz_0) \leqslant \exp(-nsz_0) \prod_{i=1}^{n} g_i(s) \qquad (3.12)$$

для всех $s \geqslant 0$, *таких, что* $g_i(s)$ *существует. Когда* z_i *дискретна, справедливо то же самое утверждение, если считать, что* $P_i(z_i)$ *— это вероятности, а интеграл, определяющий* $g_i(s)$, *заменить суммой.*

Для того чтобы применить теорему к $D(u_0 v) = \sum_{i=1}^{n} \delta(x_{i0}, y_i)$, будем рассматривать $\delta(x_{i0} y_i)$ как случайную величину, где x_i задано, а y_i имеет распределение $P_{x_i}(y_i)$. Тогда производящая функция моментов δ равна

$$g_i(s) = \int_{-\infty}^{\infty} \exp[s\delta(x_{i0} y_i)] P_{x_{i0}}(y_i) dy_i. \qquad (3.13)$$

Используя равенство (3.6), получаем

$$g_i(s) = \int_{-\infty}^{\infty} [P_{x_{i0}}(y_i)]^{1-s} [f(y_i)]^s dy_i. \qquad (3.14)$$

Для $x_{i0} = 0$ равенство (3.14) переходит в равенство

$$g_i(s) = \int_{-\infty}^{\infty} P_0(y)^{1-s} f(y)^s dy. \qquad (3.15)$$

При $x_{i0} = 1$, используя условия симметрии (3.1) и (3.5), переписываем равенство (3.14) в виде

$$g_i(s) = \int_{-\infty}^{\infty} P_0(-y)^{1-s} f(-y)^s dy. \qquad (3.16)$$

Заменяя переменную интегрирования $-y$ на y, видим, что выражения (3.15) и (3.16) идентичны, поэтому $g_i(s)$ не зависит от x_{i0} и i:

$$g_i(s) = g(s) = \int P_0(y)^{1-s} f(y)^s \, dy. \qquad (3.17)$$

Можно показать, что выражение (3.17) эквивалентно функции распределения расстояния между входом и выходом канала и не зависит от входа, что естественно ввиду симметрии канала.

Используя теорему 3.1, получим, наконец,

$$P_1 \leqslant \Pr[D(u_0 v) \geqslant nd] \leqslant g(s)^n \exp(-nsd) \qquad (3.18)$$

для всех $s \geqslant 0$, таких, что $g(s)$ в равенстве (3.17) существует.

Для того чтобы довести дело до конца, нам нужно оценить P_2, задаваемое неравенством (3.11). Для этого необходима следующая теорема, доказываемая в приложении Б.

Теорема 3.2. *Пусть z_i и w_i, $1 \leqslant i \leqslant n$ — это n пар случайных величин с плотностями распределения $P_i(z_i, w_i)$. Пусть совместная производящая функция моментов z_i, w_i задается выражением*

$$h(r, t) = \int\int \exp(rz_l + tw_l) P_l(z_l, w_l) \, dz_l \, dw_l, \qquad (3.19)$$

и пусть теперь каждая пара случайных величин не зависит от всех других пар; определим тогда Z и W следующим образом:

$$Z = \sum_{l=1}^{n} z_l, \quad W = \sum_{l=1}^{l} w_l, \quad l \leqslant n. \qquad (3.20)$$

Теперь для любых чисел z_0 и w_0

$$\Pr(Z \leqslant nz_0;\ W \leqslant nw_0) \leqslant$$

$$\leqslant \prod_{l=1}^{l} [h_l(r, t)] \prod_{l=l+1}^{n} [h_l(r, 0)] \exp[-n(rz_0 + tw_0)] \qquad (3.21)$$

для всех $r \leqslant 0$, $t \leqslant 0$, *таких, что* $h(r, t)$ *существует. Для дискретных* z *и* w *неравенство* (3.21) *по-прежнему справедливо, если интеграл в равенстве* (3.19) *заменить суммой, а плотность распределения — вероятностью.*

Теорема будет использована при оценке $\Pr[D(u_0 v) \leqslant nd;\ D(u_j, v) - D(u_0, v) \leqslant 0]$ для каждого кодового слова u_j. Допустим сначала, что u_j отличается от u_0 в первых l символах и совпадает с u_0 в последних $n - l$ символах. Тогда

$$D(u_j v) - D(u_0 v) = \sum_{i=1}^{l} \delta(x_{ij}, y_i) - \delta(x_{i0}, y_i).$$

Из условий симметрии (3.1) и (3.5) и определения δ в выражении (3.6) мы видим, что $\delta(x_{ij}, y_i)$ равно $\delta(x_{i0}, -y_i)$ при $i \leqslant l$, поскольку мы допустили, что $x_{ij} \neq x_{i0}$ для $i \leqslant l$. Пусть теперь

$$z_i = \delta(x_{i0}, y_i); \quad w_i = \delta(x_{i0}, -y) - \delta(x_{i0}, y_i).$$

Для фиксированного x_i как z_i, так и w_i являются функциями от y_i; мы можем переписать выражение для $h_i(r, t)$ в равенство (3.19) следующим образом:

$$h_i(r, t) = \int_{-\infty}^{\infty} \exp[r\delta(x_{i_0}, y_i) + t\delta(x_{i_0}, -y) - t\delta(x_{i_0}, +y_i)] P_{x_{i0}}(y_i)\, dy_i. \quad (3.22)$$

Переписав $h_i(r, t)$ так же, как это было сделано с $g_i(s)$ в равенстве (3.13), увидим, что $h_i(r, t)$ не зависит от x_{i_0} и i. Поэтому

$$h_i(r, t) = h(r, t) = \int_{-\infty}^{\infty} P_0(y)^{1-r+t} P_0(-y)^{-t} f(y)^r\, dy. \quad (3.23)$$

Применив теперь теорему 3.2, получим

$$\Pr[D(u_0, v) \leqslant nd;\ D(u_j, v) - D(u_0, v) \leqslant 0] \leqslant [h(r, t)]^l [h(r, 0)]^{n-l} e^{-nrd} \quad (3.24)$$

для всех $r \leqslant 0$, $t \leqslant 0$, если u_j и u_0 отличаются в первых l символах.

Перенумеровав n символов в блоке, заметим, наконец, что оценка (3.24) справедлива для любого из $N(l)$ кодовых слов, находящихся на расстоянии l от u_0. Отсюда

$$P_2 \leqslant \sum_{l=0}^{n} N(l)[h(r, t)]^l [h(r, 0)]^{n-l} e^{-nrd} \qquad (3.25)$$

для всех $r \leqslant 0$, $t \leqslant 0$, а $h(r, t)$ задается равенством (3.23).

В соотношении (3.25) слагаемое при $l=0$ соответствует вырожденному случаю, когда одно из оставшихся слов совпадает с u_0; $N(0)$ есть число кодовых слов с индексом, отличным от 0, но совпадающих с u_0.

Неравенства (3.25) и (3.18) оценивают P_1 и P_2; ввиду неравенства (3.9) их сумма служит оценкой вероятности ошибки декодирования по максимуму правдоподобия, когда передано фиксированное кодовое слово. Оценка выражается через кодовые расстояния $N(l)$, переходные вероятности канала $P_0(y)$ и ряд произвольных параметров s, r, t, d и $f(y)$, по которым следует произвести оптимизацию. Таким образом, комбинаторная и вероятностная стороны задачи рассмотрены, и при заданном $N(l)$ правые части соотношений (3.25) и (3.18) оцениваются, по существу, одинаково просто и при больших, и при малых длинах блоков. Проблема оптимизации, однако, ни в коей мере не тривиальна, так как уравнения трансцендентны и включают ограничения на s, r, t и $f(y)$. Одно из упрощений состоит в исключении t. Соотношение (3.25) минимизируется по t, если минимизировать $h(r, t)$, что достигается, если мы положим $\partial h(r, t)/\partial t = 0$. Тогда из равенства (3.23) получаем

$$\int_{-\infty}^{\infty} \left(\ln \frac{P_0(y)}{P_0(-y)}\right) P_0(y)^{1-r+t} P_0(-y)^{-t} f(y)^r \, dy = 0. \qquad (3.26)$$

Используя симметричность $f(y)$, заметим, что при $1-r+t=-t$ интегрируемая функция в равенстве

(3.26) антисимметрична по y и интеграл поэтому равен 0. Это действительно минимум, поскольку

$$\frac{\partial^2 h(r,t)}{\partial t^2} = \int_{-\infty}^{\infty} \left(\ln \frac{P_0(y)}{P_0(-y)}\right)^2 P_0(y)^{1-r+t} P_0(-y)^{-t} f(y)^r dy \geqslant 0.$$

Заметим, наконец, что решение

$$t = \frac{r-1}{2} \tag{3.27}$$

автоматически удовлетворяет ограничению $t \leqslant 0$ при $r \leqslant 0$.

Это упрощение позволяет переписать неравенство (3.25) так:

$$P_2 \leqslant \sum_{l=0}^{n} N(l) [h(r)]^l [g(r)]^{n-l} e^{-nrd}, \tag{3.28}$$

$$h(r) = \int_{\infty}^{\infty} [P_0(y) P_0(-y)]^{\frac{1-r}{2}} [f(y)]^r dy, \tag{3.29}$$

$$g(r) = \int_{-\infty}^{\infty} [P_0(y)]^{1-r} f(y)^r dy. \tag{3.30}$$

Эти соотношения, как видно из равенств (3.17) и (3.23), используют равенство $h(r, 0) = g(r)$.

3.5. $\bar{P}_e$ для кодов и ансамблей кодов

Рассмотрим теперь вероятность ошибки для всего кода. Пусть $N_j(l)$ есть число кодовых слов на расстоянии l от слова u_j; $0 \leqslant j \leqslant M-1$. Тогда ввиду неравенств (3.18) и (3.28) вероятность ошибки декодирования при равновероятном использовании кодовых слов выражается следующим образом:

$$\bar{P}_e \leqslant \sum_{j=1}^{M-1} \frac{1}{M} \left\{ g(s)^n e^{-nsd} + \sum_{l=0}^{n} N_j l h(r)^l g(r)^{n-l} e^{-nrd} \right\} \tag{3.31}$$

при всех $s \geqslant 0$, $r \leqslant 0$.

Определим теперь

$$\overline{N(l)} = \frac{1}{M} \sum_{j=0}^{M-1} N_j(l)$$

как среднее по j число кодовых слов на расстоянии l от u_j. Неравенство (3.31) переходит тогда в неравенство

$$\bar{P}_e \leqslant g(s)^n e^{-nsd} + \sum_{l=0}^{n} \overline{N(l)}\, h(r)^l g(r)^{n-l} e^{-nrd}$$

при всех $s \geqslant 0$, $r \leqslant 0$, (3.32)

где

$$g(s) = \int_{-\infty}^{\infty} P_0(y)^{1-s} f(y)^s \, dy, \tag{3.33}$$

$$h(r) = \int_{-\infty}^{\infty} [P_0(y) P_0(-y)]^{\frac{1-r}{2}} f(y)^r \, dy. \tag{3.34}$$

Рассмотрим теперь ансамбль кодов, описанных в гл. 2. Пусть $\overline{N(l)}$ есть среднее по ансамблю кодов значение функции $\overline{N(l)}$, определенной в первом абзаце настоящего раздела для фиксированного кода; тогда неравенство (3.32) по-прежнему имеет место, а $\bar{P}_e$ определяет теперь среднюю по ансамблю вероятность ошибки декодирования. Заметим, что по крайней мере $(1-\alpha)$-доля кодов должна иметь вероятность ошибки, не превышающую $\bar{P}_e/\alpha$. Последнее вытекает из того, что если больше чем α-доля кодов обладает вероятностью ошибки, превышающей $\bar{P}_e/\alpha$, то уже эта доля кодов вносит вклад, больший чем $\bar{P}_e$, в среднюю по ансамблю вероятность ошибки.

С оценкой (3.32) довольно трудно иметь дело, во-первых, потому, что в нее входит сумма n членов, а n может быть большим; во-вторых, потому, что она содержит ряд произвольных параметров r, s, d, $f(y)$, по которым необходимо провести оптимизацию. К сожалению, в общем случае почти ничего нельзя сделать для того, чтобы упростить оценку без ее ослабления. Однако некоторые соображения о направлении последующих упрощений и ослаблений полезно

привести перед тем, как мы непосредственно перейдем к ним. Позже будет показано, что правая часть неравенства (3.32) ведет себя примерно как экспоненциально убывающая функция длины кода и для ансамблей кодов с малой плотностью проверок, и для ансамблей равновероятных кодов с проверками на четность. Таким образом, когда изучается $\bar{P}_e$ при очень больших длинах блока и в случае исследования изменений $\bar{P}_e$ с длиной блока n, основную роль играет коэффициент при n в экспоненциальной функции. Наша задача поэтому заключается в отыскании значений d, $f(y)$, r и s, оптимизирующих этот коэффициент в экспоненте. Другие части выражения, следовательно, в процессе оптимизации во внимание приниматься не будут. Получив оценку, можно, конечно, в каждом конкретном случае вернуться обратно и получить более точный результат для неравенства (3.32), однако попытки сделать это в общем случае только усложнили бы и без того сложную ситуацию.

Допустим теперь, что функцию расстояния $\bar{N}(l)$ фиксированного кода или ансамбля кодов можно оценить выражением вида

$$\bar{N}(l) \leqslant C(\lambda, n)\, e^{nB(\lambda)}, \quad \lambda = \frac{l}{n}, \tag{3.35}$$

где $C(\lambda, n)$ должно быть сравнительно малой величиной, для того чтобы рассматриваемый далее метод оказался полезным. Выражения (2.1) и (2.19) дают такие оценки для случайного ансамбля и ансамблей кодов с малой плотностью проверок. Пусть теперь

$$C_n = \max_{\lambda} C(\lambda, n). \tag{3.36}$$

Использовав C_n вместо $C(\lambda, n)$ в неравенстве (3.35), подставив результат в неравенство (3.32), немного изменив порядок суммирования, оценив сумму n раз взятым максимальным членом, получим, наконец, что

$$\bar{P}_e \leqslant \exp n\,[\ln g(s) - sd] + \\ + n\, C_n \max_{\lambda} \exp n\,[B(\lambda) + \lambda \ln h(r) + (1-\lambda)\ln g(r) - rd] \\ \text{при всех } s \geqslant 0,\ r \leqslant 0. \tag{3.37}$$

Функции g и h по-прежнему задаются соотношениями (3.33)и (3.34). В правой части неравенства (3.37) стоят два, по существу, экспоненциальных по n слагаемых. Первое убывает с ростом d при $s>0$, а второе возрастает с ростом d при $r<0$. Поэтому, если мы выберем d так, чтобы экспоненты были равны, любое изменение d увеличило бы одну из них. Такой выбор d минимизирует коэффициент при n у большей из экспоненциальных функций. Исключив этим способом d, получим

$$\bar{P}_e \leqslant (1+nC_n)\exp\left[-n\left(\min_{\lambda} E(s, r, \lambda)\right)\right] \quad \text{при } s \geqslant 0,\ r \leqslant 0, \tag{3.38}$$

$$E(s, r, \lambda) = \frac{r}{s-r}\ln g(s) - \frac{s}{s-r}[B(\lambda)+\lambda \ln h(r)+(1-\lambda)\ln g(r)]. \tag{3.39}$$

В соответствии с определением функций g и h соотношения (3.38) и (3.39) все еще зависят от $f(y)$. В приложении Б показано, что $E(s, r, \lambda)$ максимизируется по $f(y)$ при

$$f(y) = k\left\{\frac{\left[P_0(y)^{(1-r)/2}+P_1(y)^{(1-r)/2}\right]^2}{P_0(y)^{1-s}+P_1(y)^{1-s}}\cdot + \right.$$
$$\left. + \frac{\alpha-\lambda}{\lambda(1-\alpha)}\frac{P_0(y)^{1-r}+P_1(y)^{1-r}}{P_0(y)^{1-s}+P_1(y)^{1-s}}\right\}^{\frac{1}{s-r}}, \tag{3.40}$$

где

$$\alpha = \frac{h(r)}{g(r)+h(r)}.$$

Константа k в равенстве (3.40) произвольна и исключается в оценке $\bar{P}_e$. К сожалению, это всего лишь неявное решение, поскольку α в свою очередь зависит от $f(y)$. При любых значениях s, r и λ можно найти $f(y)$, удовлетворяющее равенству (3.40), только последовательными приближениями к α. Следовательно, оптимизация выражения (3.39) трудно выполнима даже при использовании вычислительной машины. Мы вы-

берем поэтому $f(y)$ более простым способом, а именно

$$f(y)=k\left\{\frac{\left[P_0(y)^{\frac{1-r}{2}}+P_1(y)^{\frac{1-r}{2}}\right]^2}{P_0(y)^{1-s}+P_1(y)^{1-s}}\right\}^{\frac{1}{s-r}}. \qquad (3.41)$$

Для ансамбля равновероятных кодов максимизация $\bar{P}_e$ по λ приводит, как будет показано ниже, к равенству $\lambda=\alpha$, и в этом случае выражения (3.40) и (3.41) дают одинаковые результаты. Для других ансамблей изменения, вызываемые использованием равенства (3.41) вместо (3.40), приводят только к изменениям второго порядка в экспоненте $\bar{P}_e$.

Выписывая точное выражение для производящей функции моментов (выражение (3.39)) и используя равенство (3.41), получаем (см. приложение Б)

$$\bar{P}_e\leqslant(1+nC_n)\exp-nE(s,\ r), \qquad (3.42)$$

$$E(s,\ r)=\frac{s}{s-r}\beta(\alpha)-$$

$$-\ln\int_0^\infty\left(P_0^{1-s}+P_1^{1-s}\right)^{-\frac{r}{s-r}}\left(P_0^{\frac{1-r}{2}}+P_1^{\frac{1-r}{2}}\right)^{\frac{2s}{s-r}}dy, \qquad (3.43)$$

$$\alpha=\frac{\int_0^\infty\left(P_0^{1-s}+P_1^{1-s}\right)^{-\frac{r}{s-r}}\left(P_0^{\frac{1-r}{2}}+P_1^{\frac{1-r}{2}}\right)^{\frac{2r}{s-r}}2\,(P_0P_1)^{\frac{1-r}{2}}dy}{\int_0^\infty\left(P_0^{1-s}+P_1^{1-s}\right)^{-\frac{r}{s-r}}\left(P_0^{\frac{1-r}{2}}+P_1^{\frac{1-r}{2}}\right)^{\frac{2s}{s-r}}dy}, \qquad (3.44)$$

$$\beta(\alpha)=\min_\lambda\left[-B(\lambda)-\lambda\ln\alpha-(1-\lambda)\ln(-\alpha)\right]. \qquad (3.45)$$

Соотношения (3.42) и (3.45) дают общую оценку $\bar{P}_e$ через три параметра: s, r и λ. Равенство (3.45) можно использовать для исключения λ при любых фиксированных $s\geqslant 0$ и $r\leqslant 0$. Максимизация $E(s,\ r)$ по s и r не является, однако, простой задачей

и может даже приводить к ряду локальных максимумов. Эту максимизацию тем не менее можно выполнить при помощи вычислительной машины.

3.6. Вероятность ошибки для ансамбля равновероятных кодов

В качестве примера использования соотношений (3.42) и (3.45) рассмотрим частный случай ансамбля равновероятных кодов с проверками на четность, для которого из неравенства (2.1) имеем

$$B(\lambda) = -(1-R)\ln 2 - \lambda \ln \lambda - (1-\lambda)\ln(1-\lambda), \quad (3.46)$$

где $R = (\log_2 M)/n$ есть скорость кода.

Подставив равенство (3.46) в (3.45) и произведя минимизацию, увидим, что минимум находится в точке $\lambda = \alpha$ и имеет постоянную величину, не зависящую от α:

$$\beta(\alpha) = (1-R)\ln 2. \quad (3.47)$$

Поэтому и правая часть равенства (3.43) не зависит от α, что дает возможность упростить выражения (3.42) и (3.43) (см. приложение Б):

$$\bar{P}_e \leqslant (1 + nC_n)\, e^{-nE(s)},$$

$$E(s) = \frac{1}{1-s}(1-R)\ln 2 - \ln \int_0^\infty \left[P_0(y)^{1-s} + P_1(y)^{1-s}\right]^{\frac{1}{1-s}} dy \quad (3.48)$$

для любого s в области $0 \leqslant s \leqslant 1/2$.

Таким образом, $E(s)$ для любого s есть линейная функция R с наклоном $-s/(1-s)$. Рис. 3.2 иллюстрирует зависимость $E(s)$ от R, где s — параметр. Огибающая этого семейства кривых дает искомую зависимость $\bar{P}_e$ от R.

Два параметрических уравнения этой огибающей можно получить, положив частную производную $E(s)$ по s равной 0, что дает

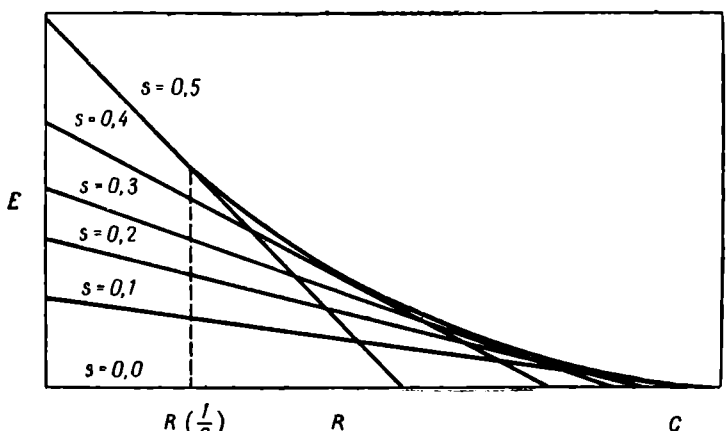
$$R(s) = 1 - \frac{(1-s)^2 \gamma'(s)}{\ln 2}, \quad 0 \leqslant s \leqslant \frac{1}{2}, \quad (3.49)$$

$$E(s) = s(1-s)\gamma'(s) - \gamma(s), \quad (3.50)$$

где

$$\gamma(s)=\ln\int_0^\infty [P_0(y)^{1-s}+P_1(y)^{1-s}]^{\frac{1}{1-s}}\,dy. \quad (3.51)$$

Можно показать следующее: $R(s)$ убывает с s, $E(s)$ возрастает с s, наклон $E(s)$ как функции от



$$E(s,R)=\frac{s}{1-s}(1-R)\ln 2-\gamma(s)$$

$$\gamma(s)=\ln\int_0^\infty \left[P_0(y)^{1-s}+P_1(y)^{1-s}\right]^{\frac{1}{1-s}}dy$$

Рис. 3.2. Семейство кривых, связывающих экспоненту и скорость в ансамбле равновероятных кодов.

$$E(s,\ R)=\frac{s}{1-s}(1-R)\ln 2-\gamma(s),$$

$$\gamma(s)=\ln\int_0^\infty [P_0(y)^{1-s}+P_1(y)^{1-s}]^{1/(1-s)}\,dy.$$

$R(s)$ равен $(-s\ln 2)/(1-s)$ и $\lim\limits_{s\to 0}R(s)$ равен пропускной способности канала. Для значений R, меньших $R(1/2)$, зависимость E от R задается соотноше-

нием (3.48) при $s=1/2$

$$E=(1-R)\ln 2-\ln\int_0^\infty [\sqrt{P_0(y)}+\sqrt{P_1(y)}]^2\,dy$$

при $R\leqslant R(1/2)$. (3.52)

Зависимость E от R, задаваемая равенствами (3.49) и (3.50), совпадает с найденной Фано [4], если не считать небольших изменений в обозначениях. Соотношения становятся еще проще в специальном случае двоичного симметричного канала (см. рис. 3.1). В этом случае

$$\gamma(s)=\frac{1}{1-s}\ln[(1-p)^{1-s}+p^{1-s}].$$

После некоторых очевидных преобразований получим знакомые результаты:

$$R(s)=1-\frac{H(p_s)}{\ln 2}, \tag{3.53}$$

$$E(s)=p_s\ln\frac{1}{p}+(1-p_s)\ln\frac{1}{1-p}-H(p_s), \tag{3.54}$$

где

$$p_s=\frac{p^{1-s}}{p^{1-s}+(1-p)^{1-s}},\quad 0\leqslant s\leqslant\frac{1}{2},$$

$$E=(1-R)\ln 2-2\ln(\sqrt{p}+\sqrt{1-p}) \tag{3.55}$$

при $R\leqslant R(1/2)$.

Мы видели, что для ансамбля равновероятных кодов значение λ, приводящее к наибольшему вкладу в $\overline{P}_e$, равно α, где α задается равенством (3.44), которое в случае равновероятного ансамбля упрощается (см. приложение Б) следующим образом:

$$\alpha(s)=\frac{\int_0^\infty (P_0^{1-s}+P_1^{1-s})^{\frac{1}{1-s}}\left[\frac{2(P_0P_1)^{1-s}}{(P_0^{1-s}+P_1^{1-s})^2}\right]dy}{\int_0^\infty (P_0^{1-s}+P_1^{1-s})^{\frac{1}{1-s}}\,dy}. \tag{3.56}$$

Любопытное следствие этого факта состоит в следующем. Предположим, что у нас есть способ увеличить минимальное расстояние типичного случайного кода. Влияние такого улучшения на вероятность ошибки декодирования в конкретном канале будет пренебрежимо малым до тех пор, пока минимальное расстояние не станет больше $n\alpha(s)$, поскольку именно на этом расстоянии возникает большая часть ошибок.

С другой стороны, если скорость кода достаточно мала, минимальное расстояние можно сделать достаточно большим для того, чтобы изменить экспоненту $\bar{P}_e$. В гл. 2 было показано, что ансамбль случайных кодов можно улучшить так, чтобы он включал только коды с минимальным расстоянием, не меньшим $n\lambda_0$, где

$$H(\lambda_0) = (1 - R) \ln 2.$$

Минимизируя $\beta(\alpha)$ в равенстве (3.45), для этого улучшенного ансамбля получаем

$$\beta(\alpha) = (1 - R) \ln 2 = H(\lambda_0), \quad \alpha > \lambda_0, \tag{3.57}$$

$$\beta(\alpha) = -\lambda_0 \ln \alpha - (1 - \lambda_0) \ln (1 - \alpha), \quad \alpha \leqslant \lambda_0. \tag{3.58}$$

Мы видим теперь, что в улучшенном ансамбле можно использовать те же значения s и r для заданной скорости, что и в неулучшенном ансамбле, и нетрудно получить экспоненциальную оценку для $\bar{P}_e$. Если мы поступим таким образом, экспонента E не изменится по сравнению со случаем неулучшенного ансамбля для скоростей, при которых $\alpha > \lambda_0$, и увеличится при $\alpha \leqslant \lambda_0$. Можно показать, что это выражение для экспоненты в самом деле максимально по s и r. Можно, кроме того, показать, что $\alpha = \lambda_0$ при некотором R_0, удовлетворяющем $0 < R_0 < R(1/2)$, и $\alpha < \lambda_0$ при $R < R_0$. При $R < R_0$ подстановка равенства (3.58) в (3.43) при $s = 1/2$, $r = 0$, и некоторые упро-

щения дают следующее:

$$E = -\lambda_0 \ln \int_0^\infty 2\sqrt{P_0(y)P_1(y)}\,dy, \qquad (3.59)$$

где λ_0 удовлетворяет уравнению $H(\lambda_0) = (1-R)\ln 2$. На рис. 3.3 показана зависимость E от R для случая

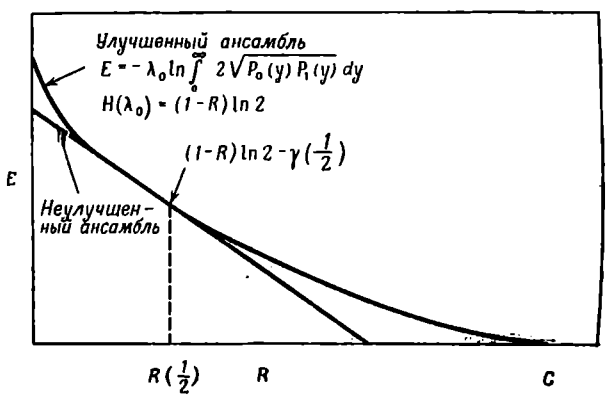


Рис. 3.3. Улучшенный и неулучшенный ансамбли равновероятных кодов.

улучшенного ансамбля. Эта оценка для двоичного симметричного канала получена ранее и независимо в еще не опубликованной работе Элайеса.

3.7. Двоичный симметричный канал

Чтобы лучше понять поведение соотношений (3.42) и (3.45) для произвольных ансамблей кодов, и в частности для ансамблей кодов с малой плотностью проверок, рассмотрим двоичный симметричный канал с вероятностью перехода p, которая приведена на рис. 3.1. Для такого канала интегралы в равенствах (3.43) и (3.44) сводятся к единственному члену, и мы

получаем

$$\bar{P}_e \leqslant (1 + nC_n) \exp[-nE(s, r)], \tag{3.60}$$

$$E(s, r) = \frac{s}{s-r}\beta(\alpha) + \frac{r}{s-r}\ln\left[(1-p)^{1-s} + p^{1-s}\right] - \\ - \frac{2s}{s-r}\ln\left[(1-p)^{\frac{1-r}{2}} + p^{\frac{1-r}{2}}\right], \tag{3.61}$$

$$\alpha = \frac{2\left[(1-p)\,p\right]^{\frac{1-r}{2}}}{\left[(1-p)^{\frac{1-r}{2}} + p^{\frac{1-r}{2}}\right]}, \tag{3.62}$$

$$\beta(\alpha) = \min_{\lambda}\left[-B(\lambda) - \lambda\ln\alpha - (1-\lambda)\ln(1-\alpha)\right]. \tag{3.63}$$

В приложении Б показано, что $E(s, r)$ имеет максимум в области $0<s<\infty$, $-\infty<r<0$ и этот максимум задается следующим образом:

$$E = \max_{s,\,r} E(s, r) = p_s \ln\frac{1}{p} + (1-p_s)\ln\frac{1}{1-p} - H(p_s), \tag{3.64}$$

где p_s — решение следующих двух уравнений с неизвестными p_s и p_r:

$$p_s = \frac{\lambda_0}{2} + (1-\lambda_0)\,p_r, \tag{3.65}$$

$$H(p_s) = B(\lambda_0) + \lambda_0\ln 2 + (1-\lambda_0)H(p_r). \tag{3.66}$$

В уравнениях (3.65) и (3.66) λ_0 есть значение λ, максимизирующее выражение

$$B(\lambda) + \frac{\lambda}{2}\ln 4p_r(1-p_r). \tag{3.67}$$

Значения s и r, при которых достигается максимум выражения (3.64), задаются неявно посредством равенств

$$\left.\begin{aligned} p_s &= \frac{p^{1-s}}{p^{1-s} + (1-p)^{1-s}}, \\ p_r &= \frac{p^{1-r}}{p^{1-r} + (1-p)^{1-r}}. \end{aligned}\right\} \tag{3.68}$$

Решение уравнений (3.65), (3.66) и (3.67) все еще требует совместного решения трех уравнений, два из

которых трансцендентны. Уравнения, однако, обладают тем преимуществом, что не содержат переходной вероятности канала p. Поэтому, если решение уравнений существует, оно справедливо для всех переходных вероятностей в области

$$p_r \leqslant p \leqslant p_s. \tag{3.69}$$

Из равенств (3.68) следует, что это та область p, для которой $s \geqslant 0$ и $r \leqslant 0$. На рис. 3.4 дается геометриче-

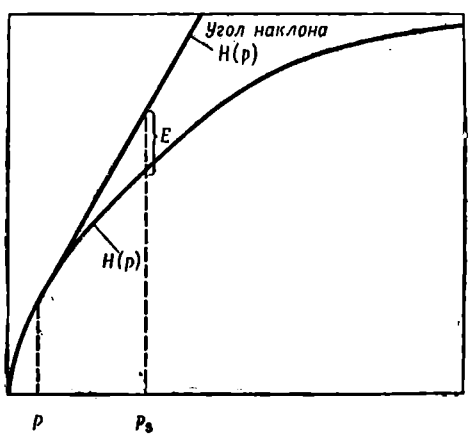
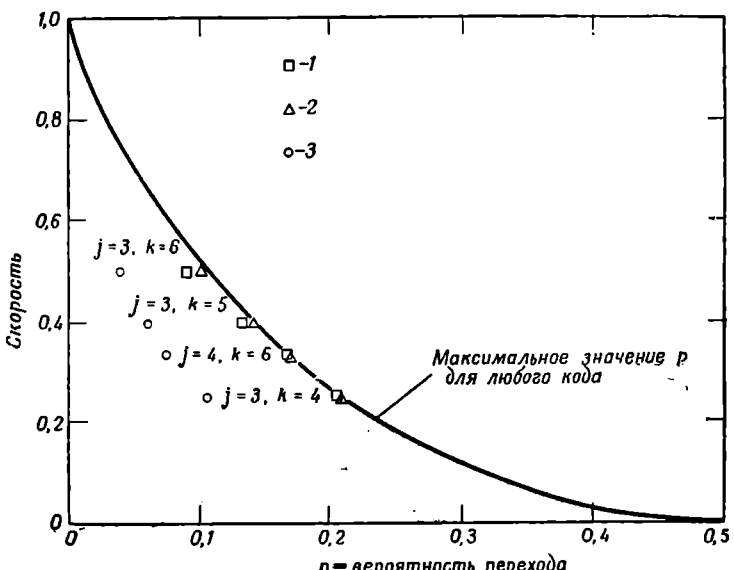


Рис. 3.4. Геометрическая интерпретация экспоненты в двоичном симметричном канале.

ская интерпретация экспоненты E, задаваемой выражением (3.64), как функции p_s и p. Интересно отметить, что выражение (3.64) совпадает с выражением (3.54), которое задает экспоненту, полученную для равновероятного ансамбля, с тем исключением, конечно, что значения p_s могут быть различными. Можно, кроме того, получить нижнюю оценку $\bar{P}_e$ для лучшего возможного кода со скоростью R; можно показать [4], что равенства (3.53) и (3.54) связывают экспоненту $\bar{P}_e$ и скорость для лучшего возможного кода. Естественно сравнивать коды в ДСК по величине параметра p_s. Было получено решение уравнений (3.65),

(3.66) и (3.67) для некоторых улучшенных ансамблей кодов с малой плотностью проверок на четность, для которых функция $B(\lambda)$ оценивается выражением (2.20). На рис. 3.5 сравниваются скорости кодов с малой плотностью проверок на четность и скорость



Р и с. 3.5. Зависимость исправляющей способности (n, j, k)-кодов от скорости в ДСК при больших n.

1. p_s — нижняя граница максимума исправляемых p при декодировании по методу максимума правдоподобия.
2. Верхняя граница максимума исправляемых p.
3. Нижняя граница максимума исправляемых p при вероятностном декодировании.

оптимального кода при одном и том же значении p_s и потому при одинаковой экспоненте $\bar{P}_e$ в области $p_r \leqslant p \leqslant p_s$. Интересно отметить, что сравнение рис. 2.4 и 3.5 указывает на то, что для этих кодов вероятность ошибки может убывать экспоненциально с ростом длины блока, даже если среднее число ошибок в блоке много больше минимального расстояния.

Таким образом, хотя ошибка декодирования и может произойти, когда число переходов в канале равно половине минимального расстояния, эта ошибка маловероятна до тех пор, пока число переходов не станет много больше минимального расстояния. Интересно также отметить, что, по-видимому, $\lambda_0 n$ представляет собой наиболее вероятное расстояние между переданным и декодированным словом в том случае, когда произошла ошибка декодирования. Более точно: это расстояние, при котором оценка вероятности ошибки максимальна. Любопытно, что эта величина не меняется при изменениях p между p_r и p_s. $E(s, r)$ для $p<p_r$ достигает максимума при $r=0$. Это и не удивительно, поскольку из равенств (3.68) следует, что $r=0$ при $p=p_r$. Подставляя $r=0$ в равенство (3.61), после некоторых алгебраических преобразований получаем

$$\max_{s,\,r} E(s,\,r) = \min_{\lambda}\left[-B(\lambda)+\frac{\lambda}{2}\ln\frac{1}{4p(1-p)}\right] \quad (3.70)$$

$$\text{при } p \leqslant p_r.$$

При $p \leqslant p_r$ значение λ, минимизирующее выражение (3.70), в типичных случаях убывает вместе с p до коэффициента минимального расстояния кода.

Предыдущие результаты получены для тех кодов и ансамблей кодов, для которых разрешимы уравнения (3.65), (3.66) и (3.67). К сожалению, они разрешимы не для всех кодов. Ни одно решение не соответствует тому случаю, когда $E(s, r)$ максимизируется при $r=-\infty$. Выражение (3.64) справедливо и в этом случае при $p \leqslant p_s$, но теперь p_s равно $\lambda_0/2$, а λ_0 есть теперь отношение минимального расстояния к длине блока. Физически это означает, что существует так много кодовых слов на минимальном расстоянии, что маловероятно исправление ошибок, если число их превышает $\lambda_0/2$. Примером может служить код всего с двумя словами, одно из которых служит дополнением другого.

3.8. Верхняя оценка скорости кодов с малой плотностью проверок на четность

Все полученные до сих пор результаты о вероятностях ошибки были оценками $\bar{P}_e$ сверху. Мы показали, что коды с малой плотностью проверок на четность ведут себя в ДСК по крайней мере так же хорошо, как и оптимальный код с несколько большей скоростью. Однако нет прямого способа показать, что некоторые коды с малой плотностью проверок не ведут себя гораздо лучше, чем средний. Некоторым шагом в этом направлении является следующая теорема, показывающая, что нельзя эффективно использовать такие коды в ДСК, пропускная способность которого сколь угодно близка к скорости кода.

Теорема 3.3. *Пусть код с проверками на четность длины n, со скоростью R и с k символами в каждом проверочном множестве используется в* ДСК *с вероятностью перехода p, и пусть кодовые слова равновероятны. Пусть*

$$H(p) = -p \ln p - (1-p)\ln(1-p),$$

$$p_k = \frac{1+(1-2p)^k}{2}.$$

Тогда при фиксированном k и при

$$R > \frac{H(p_k) - H(p)}{H(p_k)} \tag{3.71}$$

вероятность ошибки декодирования ограничена снизу величиной, не зависящей от n и большей 0.

Пропускная способность ДСК в битах на символ равна $1 - [H(p)/\ln 2]$. Поскольку $H(p_k) < \ln 2$, теорема утверждает, что для надежной передачи необходимо, чтобы скорость источника была существенно меньше пропускной способности канала. На рис. 3.5 для некоторых значений j и k показано, насколько пропускная способность должна превышать скорость источника.

Доказательство. Пусть u есть переданное кодовое слово, а v — принятая последовательность.

Тогда средняя взаимная информация в битах на символ равна

$$\frac{1}{n}\overline{I(u, v)} = \begin{cases} -\frac{1}{n}\overline{\log_2 p(u)} + \frac{1}{n}\overline{\log_2 p_v(u)}, \\ -\frac{1}{n}\overline{\log_2 p(v)} + \frac{1}{n}\overline{\log_2 p_u(v)}. \end{cases} \quad (3.72)$$

Если ненадежность[1]) информации на символ удовлетворяет неравенству

$$-\frac{1}{n}\overline{\log_2 p_v(u)} \geqslant \varepsilon > 0 \quad (3.73)$$

для некоторого ε, не зависящего от n, то вероятность ошибки декодирования также оказывается величиной, большей некоторой положительной константы[2]). Мы получим неравенство (3.73), оценивая остальные слагаемые в равенстве (3.72).

Поскольку в коде всего 2^{nR} сообщений, то

$$-\frac{1}{n}\overline{\log_2 p(u)} = R. \quad (3.74)$$

Пусть задана последовательность u, каждый символ последовательности v с вероятностью p отличен от соответствующего символа последовательности u и поэтому

$$\frac{1}{n}\overline{\log_2 p_u(v)} = \frac{-H(p)}{\ln 2}. \quad (3.75)$$

Будем задавать принятую последовательность v посредством задания сначала результатов вычисления $n(1-R)$ проверочных соотношений, а затем принятых символов в некотором множестве nR линейно независимых позиций кода. Такой способ задания v эквивалентен способу задания v указанием его символов, поскольку задание одним из этих способов позволяет вычислить его задание другим способом.

Вероятность того, что проверочное соотношение удовлетворится, равна вероятности того, что в прове-

[1]) Используется также термин «неопределенность». — *Прим. ред.*

[2]) Доказательство этого факта см. Шеннон К., Работы по теории информации и кибернетике, М., ИЛ, 1963, стр. 508.

рочное множество вошло четное число ошибок; таким образом,

$$\sum_{\substack{\text{четные}\\ l}} C_k^l p^l (1-p)^{k-l} = \frac{1+(1-2p)^k}{2}. \qquad (3.76)$$

Для проверки равенства (3.76) нужно переписать правую часть в виде

$$\frac{(1-p+p)^k + (1-p+p)^k}{2}$$

и разложить ее по формуле бинома.

Таким образом, неопределенность, связанная с каждым проверочным соотношением, равна $H(p_k)/\ln 2$ битов, где $p_k=[1+(1-2p)^k]/2$. Поскольку неопределенность, связанная с каждым информационным символом, не превышает 1 бит, а зависимости могут только уменьшить общую энтропию, имеем

$$-\frac{1}{n}\overline{\log_2 p(v)} \leqslant \frac{(1-R)H(p_k)}{\ln 2} + R. \qquad (3.77)$$

Подстановка выражений (3.74), (3.75) и (3.77) в равенство (3.72) дает следующее:

$$-\frac{1}{n}\overline{\log_2 p_v(u)} \geqslant \frac{H(p)}{\ln 2} - \frac{(1-R)H(p_k)}{\ln 2}. \qquad (3.78)$$

По предположению теоремы существует $\varepsilon>0$, удовлетворяющее равенству

$$R = \frac{H(p_k) - H(p) + \varepsilon \ln 2}{H(p_k)}. \qquad (3.79)$$

Подставляя равенство (3.79) в неравенство (3.78), получим неравенство (3.73), доказывающее теорему. Ч.Т.Д.

Глава 4

ДЕКОДИРОВАНИЕ

4.1. Введение

В гл. 3 исследовалась вероятность ошибки декодирования (n, j, k)-кодов в различных каналах с двоичным входом при декодировании по максимуму правдоподобия. Этот метод декодирования удобен, так как он минимизирует вероятность ошибки и таким образом позволяет измерить эффективность кода вне зависимости от метода декодирования. Однако практическое применение декодера, работающего по максимуму правдоподобия и действительно сравнивающего принятую последовательность со всеми кодовыми словами, выглядит не очень заманчиво. Это особенно справедливо при большой длине блока, поскольку объем кодового словаря растет экспоненциально с ростом длины блока. Было бы лучше иметь сравнительно простой (в смысле конструкции, объема памяти и числа операций) декодер, даже если он несколько увеличивает вероятность ошибки. Если же требуется меньшая вероятность ошибки, можно просто увеличить длину кодового блока.

Здесь мы опишем два метода декодирования, для которых достигается, по-видимому, разумный компромисс между сложностью метода и вероятностью ошибки декодирования. Первый метод особенно прост, но применим только в ДСК и при скоростях, много меньших пропускной способности. Второй метод, основанный на декодировании непосредственно по *апостериорным* вероятностям на выходе канала, обладает большими возможностями; однако его легче понять после знакомства с первым.

По первому методу декодер вычисляет все проверки на четность и затем изменяет все символы, со-

держащиеся больше чем в некотором фиксированном числе неудовлетворившихся проверочных соотношений. Проверки на четность вычисляются снова с использованием новых значений символов, и весь процесс повторяется до тех пор, пока не будут удовлетворены все проверочные уравнения.

Такой метод декодирования естествен, если проверочные множества невелики, поскольку их большинство либо содержит один символ, искаженный передачей, либо не содержит таких символов. Таким образом, если большинство проверочных соотношений,

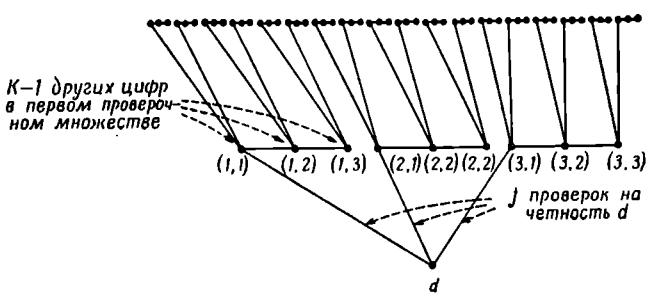


Рис. 4.1. Дерево n проверочных множеств.

проверяющих некоторый символ, оказываются неудовлетворенными, то это почти определенно указывает на то, что этот символ ошибочен. Пусть, например, при передаче исказился первый символ кода, приведенного на рис. 2.1. Тогда 1, 6 и 11 проверочные соотношения окажутся нарушенными и все три соотношения, проверяющие первый символ, не будут удовлетворены. С другой стороны, окажется неудовлетворенным не больше чем одно из трех уравнений, проверяющих любой другой символ блока.

Для того чтобы показать, как может быть исправлен некоторый символ d даже в том случае, когда его проверочное множество содержит больше одного искаженного символа, рассмотрим древовидную структуру, приведенную на рис. 4.1. Символ d представлен

корнем дерева, а каждая линия, выходящая из корня, представляет собой содержащее его проверочное множество. Все остальные символы, содержащиеся в этих проверочных множествах, представлены узлами первого яруса дерева. Линии, идущие от первого яруса ко второму, представляют собой другие проверочные множества, содержащие символы первого яруса, а узлы второго яруса представляют собой оставшиеся символы этих проверочных множеств. Заметим, что при построении следующих ярусов дерева один и тот же символ может появиться больше одного раза; такое положение мы рассмотрим в разд. 4.2.

Допустим теперь, что при передаче были искажены символ d и несколько символов первого яруса. Тогда на первом этапе декодирования неискаженные символы второго яруса и их проверочные соотношения позволяют исправить ошибки в первом ярусе. А это в свою очередь позволяет исправить символ d на втором этапе декодирования.

Итак, при декодировании некоторого символа могут оказаться полезными символы и проверочные соотношения, на первый взгляд никак не связанные с ним. Описываемый ниже вероятностный метод декодирования использует эти дополнительные символы и проверочные соотношения более систематическим образом.

4.2. Вероятностное декодирование

Допустим, что кодовые слова (n, j, k)-кода используются с равными вероятностями в произвольном канале с двоичным входом. Пользуясь обозначениями рис. 4.1, построим итерационный процесс, позволяющий вычислить для переданного символа d на m-й итерации вероятность того, что он равен 1 при условии, что известны все принятые символы вплоть до m-го яруса. На первой итерации мы можем считать, что символ d и символы первого яруса образуют подкод, в котором все множества символов, удовлетво-

ряющих j проверочным соотношениям, в дереве могут быть переданы с равными вероятностями [1]).

Рассмотрим ансамбль событий, в котором переданный символ в позиции d и все символы первого яруса суть независимые и равновероятные двоичные символы, а принятые символы в этих позициях определяются вероятностями перехода $P_x(y)$ в канале. В таком ансамбле вероятность любого события при условии, что переданные символы удовлетворяют j проверочным соотношениям, совпадает с вероятностью события в определенном выше подкоде. Так, мы хотим определить в этом ансамбле вероятность того, что переданный символ в позиции d равен 1 при условии, что известно множество принятых символов $\{y\}$ и произошло событие S, заключающееся в том, что переданные символы удовлетворяют j проверочным соотношениям, проверяющим символ d. Запишем эту вероятность следующим образом:

$$\Pr[x_d = 1 \mid \{y\}, S].$$

Пользуясь этим ансамблем и введенными обозначениями, мы можем доказать следующую теорему:

Теорема 4.1. *Пусть P_d есть вероятность того, что переданный символ в позиции d равен 1 при условии, что известен принятый символ в той же позиции, и пусть P_{il} — такая же вероятность для l-го символа i-го проверочного множества первого яруса рис. 4.1. Пусть символы независимы, а событие S состоит в том, что символы удовлетворяют j проверочным соотношениям, проверяющим символ d. Тогда*

$$\frac{\Pr[x_d = 0 \mid \{y\}, S]}{\Pr[x_d = 1 \mid \{y\}, S]} = \frac{1 - P_d}{P_d} \prod_{i=1}^{j} \left[\frac{1 + \prod_{l=1}^{k-1} (1 - 2P_{il})}{1 - \prod_{l=1}^{k-1} (1 - 2P_{il})} \right]. \quad (4.1)$$

[1]) За исключением тех случаев, когда некоторая линейная комбинация проверочных уравнений, не содержащих символа d, образует проверочное множество, состоящее из символов только первого яруса. Это положение, обсуждаемое ниже, не является, однако, серьезным ограничением.

Для доказательства нам потребуется следующая лемма.

Лемма 4.1. *Рассмотрим последовательность m независимых двоичных символов, причем вероятность того, что l-й символ последовательности равен 1, есть P_l. Тогда вероятность четного числа единиц равна*

$$1+\frac{\prod_{l=1}^{m}(1-2P_l)}{2}.$$

Доказательство леммы. Рассмотрим функцию

$$\prod_{l=1}^{m}(1-P_l+P_lt).$$

Заметим, что в разложении этого выражения по степеням t коэффициент при t^i равен вероятности появления i единиц. Функция

$$\prod_{l=1}^{m}(1-P_l-P_lt)$$

задает такое же разложение по t, за исключением того, что все коэффициенты при нечетных степенях отрицательны. При суммировании этих двух функций все коэффициенты при четных степенях удваиваются, а нечетные степени уничтожаются. И наконец, положив $t=1$ и разделив сумму на 2, получим вероятность четного числа единиц. Но

$$\frac{\prod_{l=1}^{m}(1-P_l+P_l)+\prod_{l=1}^{m}(1-P_l-P_lt)}{2}=\frac{1+\prod_{l=1}^{m}(1-2P_l)}{2},$$

что и доказывает лемму.

Доказательство теоремы.

По определению условных вероятностей имеем

$$\frac{\Pr[x_d=0\mid\{y\},S]}{\Pr[x_d=1\mid\{y\},S]}=\left(\frac{1-P_d}{P_d}\right)\left(\frac{\Pr(S\mid x_d=0,\{y\})}{\Pr(S\mid x_d=1,\{y\})}\right)^{1}). \quad (4.2)$$

При $x_d=0$ проверочное соотношение, в которое входит d, удовлетворяется, если остальные $k-1$ позиций проверочного множества содержат четное число единиц. Поскольку все символы в ансамбле независимы, вероятность того, что удовлетворены все j проверочных соотношений, равна произведению вероятностей того, что удовлетворено каждое из них. Используя лемму 4.1, получаем

$$\Pr(S\mid x_d=0,\{y\})=\prod_{l=1}^{j}\left[\frac{1+\prod_{l=1}^{k-1}(1-2P_{il})}{2}\right]. \quad (4.3)$$

Аналогично

$$\Pr(S\mid x_d=1,\{y\})=\prod_{l=1}^{j}\left[\frac{1+\prod_{l=1}^{k-1}(1-2P_{il})}{2}\right]. \quad (4.4)$$

Подставляя выражения (4.3) и (4.4) в равенство (4.2), получим утверждение теоремы. Ч.Т.Д.

Судя по сложности этого результата, может показаться трудной задача вычисления вероятности того, что переданный символ равен 1 при условии, что известны принятые символы в двух или более ярусах дерева на рис. 4.1. К счастью, однако, случай нескольких ярусов можно свести к случаю одного яруса простым итерированием.

Сначала рассмотрим случай двух ярусов. Мы можем воспользоваться теоремой 4.1 для отыскания вероятности того, что каждый переданный символ

[1]) Здесь используется то, что $P(x_d=P\{y\})=P_d$, поскольку в рассматриваемом ансамбле разные двоичные символы независимы и предполагается выполненным условие 3 определения канала в разд. 3.1. — *Прим. ред.*

первого яруса равен 1 при условии, что известны все принятые символы второго яруса. Единственное изменение, которое надо внести в теорему, состоит в том, что в первом произведении берется всего j—1 сомножителей, поскольку проверочное множество, содержащее символ d, не используется. Эти вероятности можно теперь использовать в равенстве (4.1) для отыскания вероятности того, что переданный символ в позиции d равен 1. Справедливость такого приема следует непосредственно из независимости новых значений P_{il} в ансамбле, используемом в теореме 4.1. Таким итерационным процессом можно воспользоваться по индукции для отыскания вероятности того, что переданный символ d равен 1, условной относительно любого заданного числа ярусов дерева с различными символами.

Теперь можно сформулировать общий метод декодирования для всего дерева в целом. Для каждого символа и каждой комбинации из j—1 проверочных множеств, содержащих этот символ, вычисляется, с использованием выражения (4.1), вероятность того, что передана 1, при условии, что известны принятые символы в (j—1)-м проверочном множестве. Таким образом, каждому символу соответствует j различных вероятностей, каждая из которых не учитывает одно проверочное множество. Эти вероятности затем используются в выражении (4.1) для вычисления совокупности вероятностей второго порядка. Вероятность, связываемая с некоторым символом при вычислении вероятности символа d, должна быть величиной, найденной на первой интерации, и не должна учитывать проверочное множество, содержащее символ d. При успешном декодировании вероятности, соответствующие каждому символу, стремятся (если увеличивать число итераций) либо к 1, либо к 0 (в зависимости от переданного символа). Метод справедлив только пока используются итерации, для которых не нарушается предположение о независимости в теореме 4.1. Это предположение нарушается, когда в дереве образуются петли. Поскольку каждый ярус дерева содержит в (j—1)(k—1) раз большее число

узлов, чем предыдущий, предположение о независимости должно нарушаться при сравнительно небольших m для любого кода с умеренной длиной блока. Можно, однако, пренебречь таким отсутствием независимости, сделав естественное допущение о том, что зависимости играют сравнительно небольшую роль и имеют тенденцию до некоторой степени компенсировать друг друга. Кроме того, даже если зависимости появляются на m-й итерации, первые $m-1$ итераций все же уменьшают ненадежность каждого символа. Мы можем поэтому считать, что вероятности, полученные после $m-1$ итерации, соответствуют новой принятой последовательности, декодирование которой должно оказаться проще декодирования действительно принятой последовательности.

Самое примечательное свойство этого метода заключается в том, что количество вычислений на символ и на итерацию не зависит от длины блока. Можно показать, кроме того, что среднее число итераций, требуемых для декодирования последовательности, ограничено величиной, пропорциональной логарифму логарифма длины блока.

Для практического вычисления вероятностей в теореме 4.1 оказывается удобнее преобразовать равенство (4.1) с помощью логарифмических отношений правдоподобий. Пусть

$$\left.\begin{aligned} &\ln\frac{1-P_d}{P_d}=\alpha_d\beta_d,\\ &\ln\frac{1-P_{ll}}{P_{ll}}=\alpha_{ll}\beta_{ll},\\ &\ln\left[\frac{\Pr[x_d=0\mid\{y\},S]}{\Pr[x_d=1\mid\{y\},S]}\right]=\alpha'_d\beta'_d, \end{aligned}\right\} \tag{4.5}$$

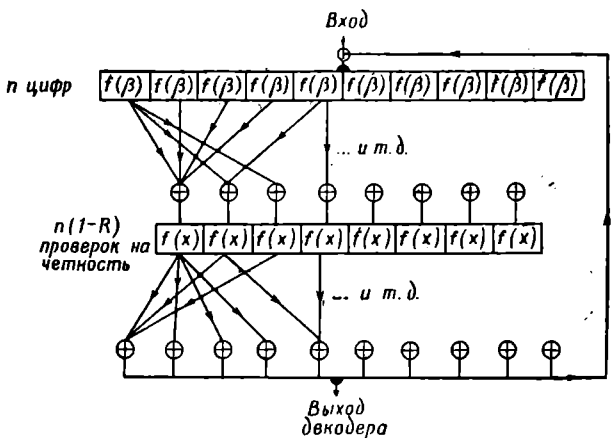
где α — знак, а β — абсолютная величина логарифмического отношения правдоподобий. После некоторых преобразований выражения (4.1) получим

$$\alpha'_d\beta'_d=\alpha_d\beta_d+\sum_{l=1}^{j}\left\{\left(\prod_{l=1}^{k-1}\alpha_{ll}\right)f\left[\sum_{l=1}^{k-1}f(\beta_{ll})\right]\right\}, \tag{4.6}$$

где

$$f(\beta)=\ln\frac{e^{\beta}+1}{e^{\beta}-1}.$$

Вычисления логарифмических отношений правдоподобий в равенстве (4.6) можно выполнять либо последовательно, либо параллельно. Последовательное вычисление можно запрограммировать на универсальной вычислительной машине; именно таким способом получены экспериментальные результаты гл. 4. Параллельное вычисление более перспективно для



Р и с. 4.2. Декодирующее устройство.

быстрого декодирования; на рис. 4.2 приведена упрощенная блок-схема, показывающая, как оно может быть выполнено.

Пусть на вход декодера подается логарифмическое отношение правдоподобий; первый ряд блоков на рис. 4.2 вычисляет для каждого символа $f(\beta)$, т. е. выполняет самую правую операцию в равенстве (4.6). На выходе сумматоров во втором ряду получается величина $\sum_{l=1}^{k-1} f(\beta_{il})$, соответствующая двум правым операциям в равенстве (4.6). Аналогично последовательные ряды блоков соответствуют операциям в равенстве (4.6), выполняемым последовательно, справа налево. На рис. 4.2, разумеется, опущены некоторые

детали, такие, как операции со знаками логарифмических отношений правдоподобий и сопоставление j различных логарифмических отношений правдоподобий каждому символу; это, конечно, не может представлять существенных трудностей.

Из схемы, представленной на рис. 4.2, видно, что вычислительное устройство параллельного действия можно легко построить, использовав аналоговые сумматоры, сумматоры по модулю 2, усилители и нелинейные устройства, аппроксимирующие функцию $f(\beta)$ в количестве, примерно пропорциональном n. Необходимая точность аппроксимации требует дальнейшего исследования; однако есть основания считать, что эта величина не является критической [1]).

4.3. Вероятность ошибки при использовании метода вероятностного декодирования

Задача математического анализа вероятностного декодирования трудна, однако можно легко получить одну очень слабую оценку вероятности ошибки.

Рассмотрим ДСК с вероятностью перехода p_0 и будем сначала исследовать (n, j, k)-код с $j=3$, в котором каждый символ содержится в трех проверочных множествах. Рассмотрим дерево проверочных множеств — такое, как приведенное на рис. 4.1 и содержащее m независимых ярусов, но перенумеруем ярусы сверху вниз, так что самый верхний ярус станет нулевым, а декодируемый символ окажется m-м ярусом.

Модифицируем метод декодирования следующим образом. Если не удовлетворены обе проверки, соответствующие ветвям, выходящим из символа первого яруса, инвертируем этот символ; проделаем то же во втором ярусе, используя измененные символы, и т. д. вплоть до символа d.

---

[1]) Недавно проведенные эксперименты показали, что в том случае, когда все вычисления выполняются в двоичной форме, достаточно иметь шесть верных значащих двоичных символов $f(\beta)$, с тем чтобы погрешность не оказывала существенного влияния на вероятность ошибки.

Вероятность ошибки декодирования символа d после проведения такого процесса служит оценкой сверху вероятности неправильного решения после m итераций в вероятностной схеме декодирования. В обоих методах решение принимается только по принятым символам в m-ярусном дереве, однако метод вероятностного декодирования всегда дает наиболее достоверное решение на основе имеющейся информации.

Найдем теперь вероятность того, что использование модифицированного метода приведет к неправильному решению относительно символа в первом ярусе. Если принятый символ искажен (вероятность такого события равна p_0), проверочные соотношения, включающие этот символ, не будут удовлетворены тогда и только тогда, когда будет искажено четное число (включая и нуль) среди остальных $k-1$ символов проверочного множества. По лемме 4.1 вероятность четного числа ошибок в $k-1$ символах равна

$$\frac{1+(1-2p_0)^{k-1}}{2}. \tag{4.7}$$

Поскольку ошибка исправляется только в том случае, когда не удовлетворены оба проверочных соотношения, содержащие символ, получим следующее выражение для вероятности того, что символ принят с ошибкой, а затем исправлен:

$$p_0\left[\frac{1+(1-2p_0)^{k-1}}{2}\right]^2. \tag{4.8}$$

Аналогичные рассуждения приводят к следующему выражению для вероятности того, что символ в первом ярусе был принят правильно, но затем был изменен из-за неудовлетворенных проверочных соотношений:

$$(1-p_0)\left[\frac{1-(1-2p_0)^{k-1}}{2}\right]^2. \tag{4.9}$$

Объединяя выражения (4.8) и (4.9), получаем вероятность неправильного решения относительно сим-

вола в первом ярусе при таком методе декодирования:

$$p_1 = p_0 - p_0\left[\frac{1+(1-2p_0)^{k-1}}{2}\right]^2 + \\ + (1-p_0)\left[\frac{1-(1-2p_0)^{k-1}}{2}\right]^2. \quad (4.10)$$

Отсюда по индукции легко следует, что если p_i есть вероятность неправильного решения после обработки символа в i-м ярусе, то

$$p_{i+1} = p_0 - p_0\left[\frac{1+(1-2p_i)^{k-1}}{2}\right]^2 + \\ + (1-p_0)\left[\frac{1-(1-2p_i)^{k-1}}{2}\right]^2. \quad (4.11)$$

Покажем теперь, что при достаточно малых p_0 последовательность $[p_i]$ сходится к 0. Рассмотрим

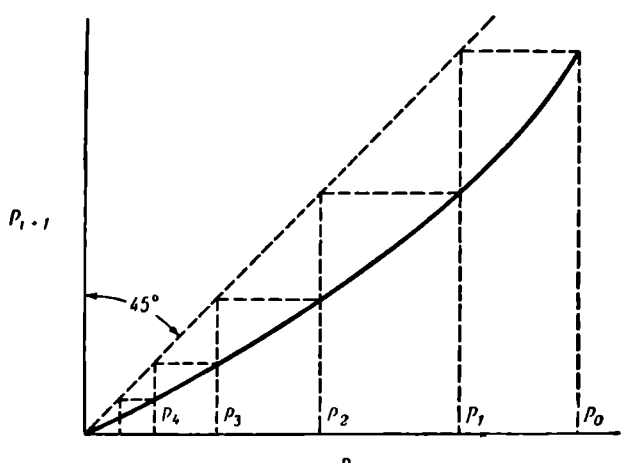


Рис. 4.3. Зависимость p_{i+1} от p_i.

рис. 4.3, на котором приведена зависимость p_{i+1} от p_i. Поскольку ордината при некотором номере i есть абсцисса для следующего, построение зигзагообразной пунктирной линии дает удобный графический метод нахождения p_i для последовательных значений i. Из

рис. 4.3 можно видеть, что если

$$\left.\begin{array}{ll} 0 < p_{i+1} < p_i & \text{для } 0 < p_i \leqslant p_0, \\ p_{i+1} = p_i & \text{для } p_i = 0, \end{array}\right\} \tag{4.12}$$

то последовательность $[p_i] \to 0$. Из равенства (4.11) видно, что неравенство в (4.12) выполняется при достаточно малых p_0. На рис. 4.4 приведены максимальные значения p_0 для нескольких значений k.

j	k	*Скорость*	p_0
3	6	0,5	0,04
3	5	0,4	0,061
4	6	0,333	0,075
3	4	0,25	0,106

Рис. 4.4. Максимальные значения p_0, при которых слабая оценка сходится.

Скорость, с которой $[p_i]$ стремится к нулю, можно определить, заметив, что из равенства (4.11) для малых p_i следует, что

$$p_{i+1} \approx p_i 2(k-1)p_0. \tag{4.13}$$

А отсюда для достаточно больших i

$$p_i \approx C[2(k-1)p_0]^i, \tag{4.14}$$

где C — константа, не зависящая от i. Поскольку число независимых итераций в дереве растет логарифмически с длиной блока, эта оценка вероятности ошибки декодирования стремится к нулю как малая отрицательная степень длины блока. Столь медленная сходимость к нулю, по-видимому, вызвана модификацией метода декодирования и требованием полной независимости и не присуща вероятностному декодированию в целом.

Подобные же рассуждения можно провести и в случае кодов, в которых число проверочных множеств на символ больше трех. Более сильный результат мо-

жно получить, если инвертировать символ всякий раз, когда не удовлетворено большее чем некоторое целое число b проверочных соотношений, содержащих этот символ; величину b мы установим ниже. Пользуясь таким критерием и рассуждая так же, как и при выводе равенства (4.11), получим

$$p_{l+1} = p_0 - \\ - p_0 \sum_{l=b}^{j-1} C_{j-1}^{l} \left[\frac{1+(1-2p_l)^{k-1}}{2}\right]^{l} \left[\frac{1-(1-2p_l)^{k-1}}{2}\right]^{j-1-l} + \\ + (1-p_0) \sum_{l=b}^{j-1} C_{j-1}^{l} \left[\frac{1-(1-2p_l)^{k-1}}{2}\right] \times \\ \times \left[\frac{1+(1-2p_l)^{k-1}}{2}\right]^{j-1-l}. \quad (4.15)$$

Теперь можно выбрать целое b так, чтобы минимизировать p_{i+1}. Решением будет наименьшее целое b, для которого

$$\frac{1-p_0}{p_0} \leqslant \left[\frac{1+(1-2p_l)^{k-1}}{1-(1-2p_l)^{k-1}}\right]^{2b-j+l}. \quad (4.16)$$

Из этого неравенства видно, что b уменьшается с уменьшением p_i. На рис. 4.5 приведена зависимость p_{i+1} от p_i с учетом того, что b выбрано меняющимся в соответствии с условием (4.16). Точка излома отвечает изменению b.

Доказательство того, что вероятность ошибки декодирования стремится к нулю с увеличением числа итераций при достаточно малых значениях вероятности перехода, остается прежним. Однако асимптотическое поведение последовательности $[p_i]$ при стремлении к нулю меняется. Из неравенства (4.16) следует, что b для достаточно малых p_i принимает значение $j/2$ при четном j и $(j+1)/2$ при нечетном j. Ис-

пользуя эти значения b и разлагая выражение (4.15) по степеням p_i, получаем

$$p_{l+1} = p_0 C_{j-1}^{\frac{j-1}{2}} (k-1)^{\frac{j-1}{2}} p_l^{\frac{j-1}{2}} + \text{члены высших порядков, } j \text{ — нечетно,} \quad (4.17)$$

$$p_{l+1} = p_0 C_{j-1}^{\frac{j}{2}} (k-1)^{\frac{j}{2}} p_l^{\frac{j}{2}} + \text{члены высших порядков, } j \text{ — четно.} \quad (4.18)$$

С помощью этих выражений можно показать, что для соответствующим образом выбранных положи-

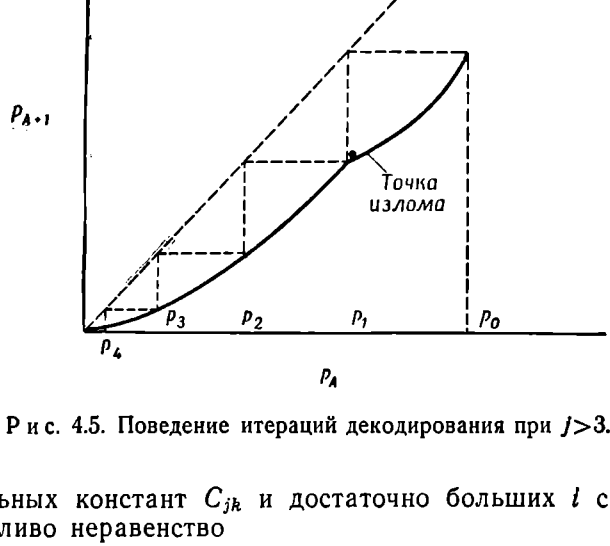


Рис. 4.5. Поведение итераций декодирования при $j>3$.

тельных констант C_{jk} и достаточно больших l справедливо неравенство

$$\begin{aligned} p_l &\leqslant \exp\left[-c_{jk}\left(\frac{j-1}{2}\right)^l\right], \quad j \text{ — нечетно,} \\ p_l &\leqslant \exp\left[-c_{jk}\left(\frac{j}{2}\right)^l\right], \quad j \text{ — четно.} \end{aligned} \quad (4.19)$$

Интересно связать этот результат с длиной кодового блока. В m-м ярусе дерева всего $(j-1)^m(k-1)^m$ символов; n должно быть не меньше этой величины, что и дает нам левую часть соотношения (4.20). С другой стороны, в приложении В приводится специальный метод построения кодов, для которых выполняется правое неравенство в соотношении (4.20)

$$\frac{\ln n}{\ln (j-1)(k-1)} \geqslant m \geqslant \frac{\ln\left[\frac{n}{2k} - \frac{n}{2j(k-1)}\right]}{2\ln (k-1)(j-1)}. \quad (4.20)$$

Объединяя неравенства (4.19) и (4.20), получаем оценку вероятности ошибки декодирования кода, удовлетворяющего условиям (4.20):

$$\begin{aligned} P_m &\leqslant \exp\left\{-c_{jk}\left[\frac{n}{2k} - \frac{n}{2j(k-1)}\right]^{\alpha}\right\}; \\ \alpha &= \frac{\ln\frac{j-1}{2}}{2\ln(j-1)(k-1)}, \quad j \text{ — нечетно}, \\ \alpha &= \frac{\ln\frac{j}{2}}{2\ln(j-1)(k-1)}, \quad j \text{ — четно}. \end{aligned} \quad (4.21)$$

При $j>3$ эта оценка вероятности ошибки декодирования убывает экспоненциально с корнем из n. Заметим, что если бы число итераций m, выполняемых без возникновения зависимостей, было в

$$[2\ln(j-1)(k-1)]/(\ln j/2)$$

раз больше, вероятность ошибки декодирования убывала бы экспоненциально с n. Существует гипотеза о том, что итерирование после проявления зависимостей в вероятностной схеме декодирования даст такую экспоненциальную оценку.

Другой способ оценки вероятностного метода декодирования состоит в вычислении распределения вероятностей логарифмических отношений правдоподобий для ряда итераций. Такой подход позволяет выяснить, можно ли с помощью кода с фиксированными

j и k добиться сколь угодно малой вероятности ошибки при передаче по заданному каналу. Расчеты на вычислительной машине ИБМ-709 показали, что код с $j=3$ и $k=6$ пригоден для каналов с вероятностями перехода вплоть до 0,07, а код с $j=3$ и $j=4$ пригоден для каналов с вероятностями вплоть до 0,144. Эти цифры особенно интересны, поскольку опровергают установившееся мнение о том, что пороговая скорость последовательного декодирования ограничивает интервал скоростей, при которых вообще возможны сколько-нибудь простые методы декодирования.

Глава 5

КОДЫ С МАЛОЙ ПЛОТНОСТЬЮ ПРОВЕРОК НА ЧЕТНОСТЬ С АЛФАВИТОМ ПРОИЗВОЛЬНОГО ОБЪЕМА

Результаты гл. 2, 3 и 4 о двоичных кодах с малой плотностью проверок на четность мы распространим в этой главе на коды с алфавитом произвольного объема. Буквами алфавита будут A-ичные символы, где A — объем алфавита; A-ичные символы суть числа от 0 до $A-1$ включительно. Определения (n, j, k)-матриц и ансамблей матриц совпадают с определениями гл. 2. Кодовые слова, задаваемые такой матрицей, суть последовательности A-ичных символов, такие, что сумма символов, входящих в проверочное множество, равна 0 по модулю A.

5.1. Функции расстояния

Определим расстояние между двумя последовательностями в коде с алфавитом объема A как число позиций, в которых последовательности различаются. Вес последовательности равен числу символов, отличных от нуля, т. е. расстоянию от нулевой последовательности. Функция расстояния кода $N(l)$ определяется снова как число кодовых слов веса l. Из групповых свойств такого кода следует [12], что $N(l)$ — число кодовых слов на расстоянии l от любого кодового слова. Для того чтобы получить верхнюю оценку $N(l)$ этих кодов, нам потребуется следующая теорема, являющаяся прямым обобщением теоремы 2.3.

Теорема 5.1. *Для каждого кода в (n, j, k)-ансамбле с алфавитом объема A число $N_1(l)$ последовательностей веса l, удовлетворяющих любому из j*

блоков по n/k проверок, оценивается следующим образом:

$$N_1\left[\frac{n}{k}\mu'_A(s)\right] \leqslant \exp\frac{n}{k}\left[\mu_A(s) - s\mu'_A(s) + (k-1)\ln A\right], \quad (5.1)$$

где s есть произвольный параметр, а $\mu_A(s)$ определяется следующим образом:

$$\left.\begin{aligned} \mu_A(s) &= \ln A^{-k}\left\{[1+(A-1)e^s]^k + \right. \\ &\qquad \left. + (A-1)(1-e^s)^k\right\}, \\ \mu'_A(s) &= \frac{d\mu_A(s)}{ds}. \end{aligned}\right\} \quad (5.2)$$

Доказательство. Рассмотрим фиксированное проверочное множество из k символов. Пусть $m(l)$ есть число различных последовательностей A-ичных символов длины k, веса l и таких, что сумма компонент каждой равна 0 по модулю A. Покажем сначала, что для произвольного t справедливо равенство

$$\sum_{l=0}^{k} m(l)\, t^l = \frac{1}{A}[1+(A-1)t]^k + \frac{A-1}{A}(1-t)^k. \quad (5.3)$$

Рассмотрим двойную перечисляющую функцию

$$B(t, r) = (1 + \mathrm{tr} + \mathrm{tr}^2 + \ldots + \mathrm{tr}^{A-1})^k = \quad (5.4)$$

$$= \sum_{l,\,j} b_{lj} t^l r^j. \quad (5.5)$$

Ясно, что b_{lj} есть число последовательностей длины k, содержащее l ненулевых A-ичных символов, таких, что их сумма равна j. Теперь рассмотрим выражение

$$\frac{1}{A}\sum_{a=0}^{A-1} B\left(t, e^{\frac{i2\pi a}{A}}\right) = \sum_{l,\,j} b_{lj} t^l \left(\frac{1}{A}\sum_{a=0}^{A-1} \exp\frac{ij2\pi a}{A}\right). \quad (5.6)$$

Сумма в скобках в равенстве (5.6) равна 0 при всех j, не кратных A, ввиду равномерного распределения слагаемых по единичному кругу комплексной плоско-

сти. Если j кратно A, сумма в скобках равна 1. Таким образом,

$$\frac{1}{A}\sum_{a=0}^{A-1} B\left(t,\ e^{\frac{i2\pi a}{A}}\right)=\sum_{l=0}^{k} m(l)\,t^l. \tag{5.7}$$

И наконец, для $r\neq 1$ из выражения (5.4) получаем

$$B(t,\ r)=\left[1+t\left(\frac{r-r^A}{1-r}\right)\right]^k; \tag{5.8}$$

$$B\left(t,\ e^{\frac{i2\pi a}{A}}\right)\begin{cases}=(1-t)^k, & a\neq 0,\\ =[1+(A-1)\,t]^k, & a=0.\end{cases} \tag{5.9}$$

Объединяя равенства (5.9) и (5.7), получаем равенство (5.3).

Рассмотрим теперь ансамбль, в котором все A-ичные последовательности длины n и удовлетворяющие заданным n/k проверочным соотношениям равновероятны. Тогда для любых k символов в проверочном множестве все A^{k-1} последовательностей длины n, удовлетворяющих проверочному соотношению, также равновероятны; из равенства (6.3) получаем поэтому следующее выражение для производящей функции моментов весов этих последовательностей:

$$g(s)=A^{-k}[1+(A-1)e^s]^k+(A-1)(1-e^s)^k. \tag{5.10}$$

Утверждение теоремы теперь получается точно так же, как и в теореме 2.3.

Всего существует $C_n^l(A-1)^l$ A-ичных последовательностей веса l; таким образом, вероятность того, что выбранная случайным образом последовательность веса l удовлетворит блоку из n/k проверочных соотношений, равна

$$\frac{N_1(l)}{C_n^l(A-1)^l};$$

Поскольку в ансамбле кодов все j блоков проверочных соотношений независимы, вероятность того,

что последовательность веса l будет кодовым словом, равна

$$P(l)=\left[\frac{N_1(l)}{C_n^l(A-1)^l}\right]^j.$$

Отсюда, следуя построениям гл. 2, можно оценить функцию расстояния $\overline{N_{jk}(l)}$ и функцию распределения минимального расстояния

$$\overline{N_{jk}(\lambda n)}\leqslant C(\lambda,\ n)\exp[-nB_{jkA}(\lambda)], \qquad (5.11)$$

$$\Pr(D\leqslant n\delta)\leqslant\sum_{l=2}^{n\delta}C(\lambda,\ n)\exp\left[-nB_{jkA}\left(\frac{l}{n}\right)\right]. \qquad (5.12)$$

где

$$B_{jkA}(\lambda)=(j-1)[H(\lambda)+\lambda\ln(A-1)]-$$
$$-\frac{j}{k}[\mu_A(s)+(k-1)\ln A]+js\lambda, \qquad (5.13)$$

$$C(\lambda,\ n)=[2\pi n\lambda(1-\lambda)]^{\frac{j-1}{2}}\exp\frac{j-1}{12n\lambda(1-\lambda)}; \qquad (5.14)$$

$$\lambda=\frac{\mu_A'(s)}{k}, \qquad (5.15)$$

а $\mu_A(s)$ задается равенством (5.2).

Методом, аналогичным примененному в приложении А, можно показать, что функция $B_{jkA}(\lambda)$ равна нулю при $\lambda=0$; затем она растет вблизи нуля, причем производная в нуле бесконечна; она пересекает ось абсцисс в единственной точке, характеризующей типичное минимальное расстояние, и затем остается отрицательной.

5.2. Вероятность ошибки декодирования

Рассмотрим канал с входным алфавитом из A букв, которые мы для удобства перенумеруем A-ичными символами. Обозначим выход через y, и пусть,

как и в гл. 3, $f(y)$ есть некоторая произвольная функция, заданная на выходном алфавите. Пусть u_0, $u_1, \ldots, u_j, \ldots, u_{M-1}$ — суть M кодовых слов A-ичного блокового кода длины n и $u_j = x_{1j}, x_{2j}, \ldots, x_{nj}$. Теперь определим расстояние между входным словом $u = (x_1, \ldots, x_n)$ и выходом $v = (y_1, \ldots y_n)$ следующим образом:

$$D(u, v) = \sum_{l=1}^{n} \delta(x_l, y_l), \tag{5.16}$$

где

$$\delta(x, y) = \ln \frac{p(y \mid x)}{f(y)}. \tag{5.17}$$

Определим

$$g_l(s) = \overline{\exp s\delta_l}, \tag{5.18}$$

$$h_{lj}(r, t) = \overline{\exp r\delta_l + t(\delta_j - \delta_l)}. \tag{5.19}$$

Усреднение в равенствах (5.18) и (5.19) проведено в соответствии с условным распределением выхода канала y при переданном x_i. Ограничим свое внимание каналами, симметричными в том смысле, что $g_i(s)$ и $h_{ij}(r, t)$ не зависят от i и j при подходящем выборе $f(y)$. В дальнейшем мы будем рассматривать только те функции $f(y)$, для которых выполнено это условие симметрии.

Примером такого канала может служить канал с A-ичными символами на входе и выходе, для которого вероятность принять переданный символ равна $1 - p$, а вероятность принять некоторый отличный от него символ равна $p/1 - p$. Другим примером служат A ортогональных сигналов равной энергии либо в канале с белым гауссовским шумом, либо в канале с релеевскими замираниями, аналогичным приведенным на рис. 3.1.

Декодирование по методу максимума правдоподобия в таком канале, когда принято слово v, эквивалентно выбору u_j, минимизирующего $D(u_i, v)$. Таким образом, можно оценить вероятность ошибки декоди-

рования по методу максимума правдоподобия, когда передано u_0, следующим образом:

$$P(e) \leqslant P_1 + P_2, \tag{5.20}$$

$$P_1 \leqslant \Pr\left[\sum_{l=1}^{n} \delta(x_{l0}, y_l) \geqslant n\delta\right]; \tag{5.21}$$

$$P_2 \leqslant \sum_{j=1}^{M-1} \Pr\left[\sum_{l=1}^{n} \delta(x_{l0}, y_l) < n\delta;\ \sum_{l=1}^{n} \delta(x_{lj}, y_l) - \right.$$
$$\left. - \delta(x_{l0}, y_l) < 0\right]. \tag{5.22}$$

Тогда теоремы 3.1 и 3.2 можно непосредственно использовать для оценки в неравенствах (5.21) и (5.22):

$$P_1 \leqslant [g(s)]^n \exp(-nsd), \quad s \geqslant 0, \tag{5.23}$$

$$P_2 \leqslant \sum_{l=0}^{n} N(l)[h(r, t)]^l [h(r, 0)]^{n-l} \exp(-nrd), \tag{5,24}$$
$$r \leqslant 0, \quad t \leqslant 0,$$

где $g(s)$ и $h(r, t)$ задаются равенствами (5.18) и (5.19), а $N(l)$ есть функция расстояния кода. Для ансамбля кодов с проверками на четность правые части неравенств (5.20), (5.23) и (5.24) оценивают среднюю по ансамблю вероятность ошибки через произвольные параметры d, $f(y)$, $s \geqslant 0$, $r \leqslant 0$, $t \leqslant 0$. Как и в гл. 3, значение $t = (r-1)/2$ оптимизирует оценку по t, однако никаких других упрощений не найдено. Неравенства (5.20), (5.23) и (5.24) вместе с неравенством (5.11) позволяют тем не менее показать, что вероятность ошибки для улучшенного ансамбля (n, j, k)-кодов при достаточно малых скоростях убывает экспоненциально с ростом длины блока.

5.3. Вероятностное декодирование

Рассмотрим A-ичный (n, j, k)-код и допустим, что все кодовые слова равновероятны. Как и в гл. 4, мы хотим, используя обозначения рис. 4.1, найти $P_m(xd = a)$ — вероятность того, что переданный символ в позиции d равен a, $0 \leqslant a \leqslant A - 1$ при условии, что из-

вестны принятые символы в m ярусах проверочного дерева с корнем x_d. Найдем сначала $P_1(x_d=a)$.

Рассмотрим ансамбль, в котором переданный символ в позиции d и узлы первого яруса суть независимые равновероятные A-ичные символы, а принятый символ определяется каналом. В таком ансамбле вероятность любого события при условии, что j проверочных соотношений первого яруса удовлетворены, совпадает с вероятностью этого события в рассматриваемом ансамбле кодов. Поэтому в используемых ранее обозначениях

$$P_1(x_d=a)=\Pr(x_d=a|\{y\},\ S). \tag{5.25}$$

Теорема 5.2. *Пусть $P_0(x_{il}=a)$ есть вероятность того, что l-й переданный A-ичный символ в i-м проверочном множестве, проверяющем символ d, равен a при условии, что известен принятый символ в этой позиции. Пусть все комбинации x_d и x_{il}, удовлетворяющие j проверочным соотношениям, равновероятны. Тогда*

$$P_1(x_d=a)=\frac{P_0(x_d=a)\prod\limits_{l=1}^{j} g_l(-a)}{\sum\limits_{a=0}^{A-1} P_0(x_d=a)\prod\limits_{l=1}^{j} g_i(-a)}, \tag{5.26}$$

где

$$G_i(t)=\sum_{a=0}^{A-1} g_i(a)\,t^a=\prod_{l=1}^{k-1}\sum_{a=0}^{A-1} P_0(x_{il}=a)\,t^a. \tag{5.27}$$

В равенстве (5.26) *a берется по модулю A, а умножение в соотношении* (5.27) *выполняется по модулю t^A.*

Равенство (5.27) дает явное выражение для $g_i(a)$ при каждом a, но при вычислениях мы находим $g_i(a)$ сразу для всех a, $0\leqslant a\leqslant A-1$. Для доказательства теоремы нам потребуется следующая лемма.

Лемма 5.1. *Рассмотрим последовательность L независимых A-ичных символов, в которой l-й символ имеет распределение вероятностей $P_l(a)$. Вероятность*

того, что сумма символов по модулю A примет значение a, равна $g(a)$ в разложении

$$G(t)=\sum_{a=0}^{A-1} g(a)\,t^a=\prod_{l=1}^{L}\sum_{a=0}^{A-1} P_l(a)\,t^a, \qquad (5.28)$$

где произведение в соотношении (5.28) *берется по модулю* t^A.

Доказательство леммы. Заметим, что правая часть соотношения (5.28) при использовании обычного умножения оказывается просто z-преобразованием суммы L букв. Другими словами, коэффициент при t^a в разложении (5.28) равен вероятности того, что сумма L символов равна a. Изменения, связанные с тем, что мы берем произведение по модулю t^A, сводятся просто к тому, что коэффициенты при степенях с показателями, равными по модулю A, складываются, что и доказывает лемму.

Доказательство теоремы. После некоторых преобразований условных вероятностей из равенства (5.25) получим

$$P_1(x_l=a)=\frac{\Pr(S\mid x_d=a,\{y\})P_0(x_d=a)}{\Pr(S\mid\{y\})}=$$

$$=\frac{\Pr(S\mid x_d=a,\{y\})P_0(x_d=a)}{\sum_{a'=0}^{A-1}\Pr(S\mid x_d=a',\{y\})P_0(x_d=a')}. \qquad (5.29)$$

Заметим теперь, что $\Pr\{S\mid x_d=a,\{y\}\}$ есть вероятность того, что суммы каждой совокупности $k-1$ символов, отличных от d в проверочном множестве, равны $-a$. Из леммы 5.1 имеем

$$\Pr(S\mid x_d=a,\ \{y\})=\prod_{l=1}^{J} g_l(-a \bmod A), \qquad (5.30)$$

где $g_i(-a \bmod A)$ задается в уравнении (5.27). Подставляя выражение (5.30) в равенство (5.29), получаем утверждение теоремы. Ч.Т.Д.

Выражение (5.26) можно сразу же использовать для построения итерационного процесса, воспользовавшись рассуждениями гл. 4. При последовательных итерациях следует писать $P_m(x_{il}=a)$ вместо $P_0(x_{il}=a)$; для каждого символа вычисляются j различных вероятностей, каждая из которых не учитывает одно проверочное множество.

5.4. Вероятность ошибки при вероятностном декодировании

Рассмотрим канал с A входами и выходами, перенумерованными от 0 до $A-1$. Переходные вероятности канала задаются следующим образом:

$$p(y_a|x_a)=1-p_0; \quad p(y_a|x_b)=\frac{p_0}{A-1}$$

для всех a и b, где $a \neq b$.

Рассмотрим дерево проверочных множеств, такое, как приведенное на рис. 4.1, с m независимыми ярусами и $j=3$. Модифицируем метод декодирования следующим образом. Если оба проверочных соотношения, содержащие некоторый символ, не удовлетворены и принимают равные значения, изменим символ так, чтобы оба проверочных соотношения удовлетворились; во всех остальных случаях символ изменять не будем. Вероятность ошибки при таком методе служит оценкой сверху для вероятности ошибки при вероятностном декодировании. Вероятность того, что символ первого яруса принят искаженным, а затем исправлен, равна p_0Q^2, где Q есть вероятность того, что в одном из множеств по $k-1$ символов нет искаженных символов, или сумма искаженных символов по модулю A равна 0. Определим ошибку в символе как $(y-x) \bmod A$. Покажем теперь, что

$$Q=\frac{1+(A-1)\left(1-\frac{Ap_0}{A-1}\right)^{k-1}}{A}. \qquad (5.31)$$

Заметим, что z-преобразование обычной суммы ошибок в $(k-1)$ символах имеет следующий вид:

$$G(z)=\left(1-p_0+\frac{p_0}{A-1}\sum_{a=1}^{A-1}z^a\right)^{k-1}, \tag{5.32}$$

$$G(z)=\begin{cases}\left(1-p_0+\dfrac{p_0}{A-1}\dfrac{z-z^A}{1-z}\right)^{k-1}, & z\neq 1, \quad (5.33)\\ 1, & z=1. \quad (5.34)\end{cases}$$

Рассмотрим теперь величину

$$\frac{1}{A}\sum_{a=0}^{A-1}G(e^{j2\pi a/A}).$$

Все степени z этого выражения, не кратные A, взаимно уничтожаются ввиду их равномерного распределения по единичному кругу в комплексной плоскости, коэффициенты при степенях, кратных A, складываются, что и дает выражение для Q:

$$Q=\frac{1}{A}\sum_{a=0}^{A-1}G(e^{j2\pi a/A}). \tag{5.35}$$

Теперь, используя равенство (5.33), получаем

$$G(e^{j2\pi a/A})=\left(1-p_0-\frac{p_0}{A-1}\right)^{k-1} \quad \text{при} \quad a\neq 0. \tag{5.36}$$

И наконец, объединяя равенства (5.36) и (5.34), получаем (5.31). Таким образом, вероятность того, что символ первого яруса принят неправильно, а затем исправлен, равна

$$p_0\left\{\frac{1+(A-1)\left(1-\dfrac{Ap_0}{A-1}\right)^{k-1}}{A}\right\}^2. \tag{5.37}$$

Вероятность того, что символ первого яруса принят правильно, но изменен из-за двух одинаково искаженных проверок, равна

$$(1-p_0)(1-Q)\left(\frac{1-Q}{A-1}\right). \tag{5.38}$$

В равенстве (5.38) множитель $1-Q$ есть вероятность того, что искаженные символы одного из мно-

жеств $k-1$ символов не удовлетворяют уравнению, а множитель $(1-Q)/(A-1)$ равен вероятности того, что сумма символов второго множества принимает значение, сравнимое по модулю A с суммой первого множества.

Объединяя соотношения (5.37), (5.38) и (5.31), получаем вероятность ошибки решения о символе в первом ярусе после первой итерации процесса декодирования:

$$p_1 = p_0 - p_0\left\{\frac{1+(A-1)\left(1-\frac{Ap_0}{A-1}\right)^{k-1}}{A}\right\}^2 + \\ +(1-p_0)(A-1)\left\{\frac{1-\left(1-\frac{Ap_0}{A-1}\right)^{k-1}}{A}\right\}^2. \quad (5.39)$$

Аналогичным образом для последовательных ярусов получаем

$$p_{l+1} = p_0 - p_0\left\{\frac{1+(A-1)\left(1-\frac{Ap_l}{A-1}\right)^{k-1}}{A}\right\}^2 + \\ +(1-p_0)(A-1)\left\{\frac{1-\left(1-\frac{Ap_l}{A-1}\right)^{k-1}}{A}\right\}^2. \quad (5.40)$$

Скорость, с которой $[p_i]$ стремится к нулю, можно определить из равенства (5.40). При малых p_i

$$p_{l+1} \approx p_l 2(k-1)p_0. \quad (5.41)$$

Интересно отметить, что выражение (5.41) совпадает с выражением (4.14), но, конечно, максимальное значение p_0, при котором последовательность p_i сходится к нулю, здесь другое. С ростом A эта величина растет до $1/2(k-1)$.

Значительно труднее получить оценку вероятности ошибки декодирования для случаев $j>3$. Декодируемый символ должен изменяться всякий раз, когда b (b будет определено позже) или большее число проверочных отношений, содержащих этот символ, все принимают одно значение. Символ следует изменить таким образом, чтобы были удовлетворены все b

проверочных соотношений. При $b>(j-1)/2$ можно таким же способом, как это было сделано в разд. 4.3, показать, что

$$p_{i+1}=p_0-p_0\sum_{l=b}^{j-1}C_{j-1}^{l}Q_i^l(1-Q_i)^{j-1-l}+$$

$$+(1-p_0)\sum_{l=b}^{j-1}C_{j-1}^{l}\left(\frac{1-Q_i}{A-1}\right)^{l}\left(1-\frac{1-Q_i}{A-1}\right)^{j-1-l}(A-1), \tag{5.42}$$

где

$$Q_i=\frac{1+(A-1)\left(1+\frac{Ap_i}{A-1}\right)^{k-1}}{A}. \tag{5.43}$$

Теперь можно выбрать целое b так, чтобы минимизировать p_{i+1}, причем на b накладывается ограничение $b>(j-1)/2$. Решением задачи будет минимальное целое b, такое, что

$$\frac{1-p_0}{p_0}\leqslant\frac{Q_i^b(A-1)^{j-2}}{(1-Q_i)^{2b+1-j}(A-2-Q_i)^{j-1-l}}. \tag{5.44}$$

При p_i, стремящемся к нулю, $b=j/2$ для четных j и $b=(j+1)/2$ для нечетных j. Раскладывая выражение (5.42) по степеням p_i, получаем

$$p_{i+1}=p_0C_{j-1}^{(j-1)/2}[p_i(k-1)]^{(j-1)/2}+\ldots, \quad j-\text{нечетное}; \tag{5.45}$$

$$p_{i+1}=\left[p_0+\frac{1-p_0}{(A-1)^b}\right]C_{j-1}^{j/2}[p_i(k-1)]^{j/2}+\ldots, \quad j-\text{четное}. \tag{5.46}$$

Заметим, что выражение (5.45) совпадает с (4.17), а (5.46) — с (4.18) с точностью до коэффициентов.

При выводе вероятности ошибки из равенства (4.18) в гл. 4 мы не использовали ограничения, связанные с тем, что алфавит двоичный, поэтому оценка вероятности ошибки, задаваемая неравенством (4.21), справедлива для кодов с произвольным объемом алфавита. Однако коэффициенты c_{jk}, используемые в выражении (4.21), зависят от объема алфавита A.

Глава 6

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Вероятность ошибки декодирования $P(e)$ при некотором методе кодирования и декодирования можно измерить непосредственно, моделируя рассматриваемые канал и схему на вычислительной машине. К сожалению, для более или менее точной оценки значения $P(e)$ необходимо получить довольно большое число случаев неправильного декодирования, поэтому число повторений эксперимента должно быть много больше $1/P(e)$. При длине блока около 500 вычислительная машина ИБМ-7090 затрачивает примерно 0,1 *сек* на итерацию при декодировании блока согласно вероятностной схеме декодирования. Следовательно, для оценки вероятности $P(e)$, даже имеющей порядок 10^{-4}, потребуются многие часы машинного времени.

Ввиду ограниченности машинного времени все приведенные здесь результаты относятся к случаям больших $P(e)$. Безусловно, интереснее оказались бы результаты для малых значений $P(e)$. Однако приведенные ниже данные можно, по-видимому, с некоторой долей уверенности экстраполировать на случаи, когда $P(e)$ имеет порядок 10^{-5} или 10^{-6}. Кроме того, даже приведенные здесь весьма ограниченные результаты дают некоторую информацию о поведении $P(e)$ при изменении таких параметров, как длина блока, скорость кода и тип канала.

6.1. Моделирование кода

Все результаты этой главы получены для кодов с малой плотностью проверок на четность, моделированных на вычислительной машине ИБМ-7090 псевдо-

случайным методом. Говоря более конкретно, проверочные матрицы выбирались так же, как строился в гл. 2 ансамбль матриц с малой плотностью проверок. Первая подматрица из n/k проверочных множеств содержала по k символов, расположенных в ступенчатом порядке, и каждая из последующих подматриц строилась случайной перестановкой столбцов первой. Случайные перестановки выполнялись обычным получением псевдослучайных чисел, результаты затем изменялись так, чтобы никакие два проверочных множества не содержали больше одного общего символа. Такая модификация случайного выбора гарантирует выполнение введенных выше предположений для первой итерации, а также исключает маловероятную возможность выбрать код с минимальным расстоянием 2.

Коды, построенные таким способом, запоминались вычислительной машиной и использовались при декодировании шумовых последовательностей, имитирующих двоичный симметричный канал, канал с белым гауссовским шумом и канал с релеевскими замираниями. Для сокращения времени вычислений было принято, что передается кодовое слово, целиком состоящее из нулей. Это допустимо, поскольку в гл. 3 было показано, что вероятность ошибки декодирования при передаче по симметричному каналу с двоичным входом не зависит от передаваемого слова. Такое упрощение, конечно, требует особенно тщательной проверки полной симметрии положительных и отрицательных выходов при моделировании декодирования.

6.2. Двоичный симметричный канал

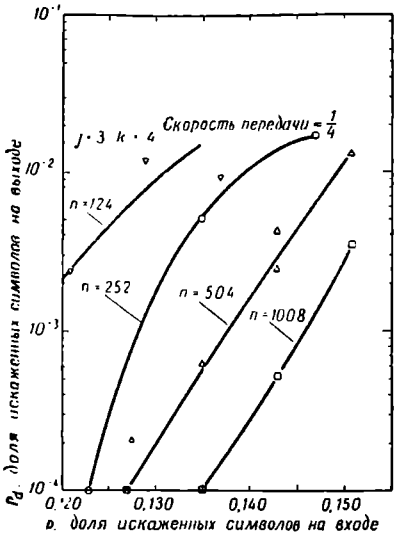
Точное моделирование двоичного симметричного канала (ДСК) состоит в выборе последовательности случайных ошибок, в которой переходы (т. е. ошибки, представляемые перекрещивающимися линиями перехода на рис. 3.1, *в*) возникают независимо и с фиксированной вероятностью p. Возможности декодирования такой последовательности при использовании вероятностного метода с конкретным кодом очень

сильно зависят от c — числа переходов, происшедших в процессе моделирования. Поскольку c подчиняется известному (биномиальному) распределению, можно при фиксированном c экспериментально оценить вероятность ошибки декодирования, а затем по этим данным вычислить $P(e)$ для ДСК. Преимущества последнего приема заключаются в том, что он позволяет получить дополнительные сведения о эффективности метода декодирования и облегчает сравнение с другими методами, рассчитанными на исправление фиксированного числа ошибок.

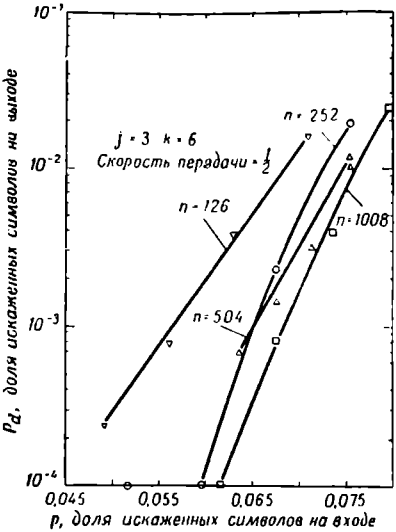
На рис. 6.1—6.4 приведены результаты, полученные таким способом. По оси абсцисс на каждом из графиков отложено c/n — отношение числа переходов к длине блока, а по оси ординат — вероятность ошибки на символ после декодирования. В экспериментах для всех кодов, за исключением кода со скоростью $1/2$ и длиной блока 126, декодер просто отказывал; он не выполнял декодирование даже неправильно. Другими словами, вычисляемая декодером *апостериорная* вероятность не сходилась ни к 1, ни к 0. А это очень важно в любой системе с каналом обратной связи, поскольку недекодированный блок информационных символов может быть переспрошен. Важно также отметить, что приведенные на рис. 6.1—6.4 значения $P(e)$ суть вероятности ошибки на символ, полученные после того, когда сделаны наилучшие возможные предсказания относительно каждого символа в блоках, которые не удалось декодировать. Вероятность неудачи при декодировании блока, как правило, превышает $P(e)$ в 10 раз.

Медиана числа блоков с отказами декодирования для точек, отложенных на графиках рис. 6.1—6.4, равна 8, значительное число точек, особенно при малых значениях $P(e)$, оценивалось при отказе только в одном-двух блоках. Поэтому весьма вероятно, что положение отдельных точек при большем объеме данных заметно изменилось бы.

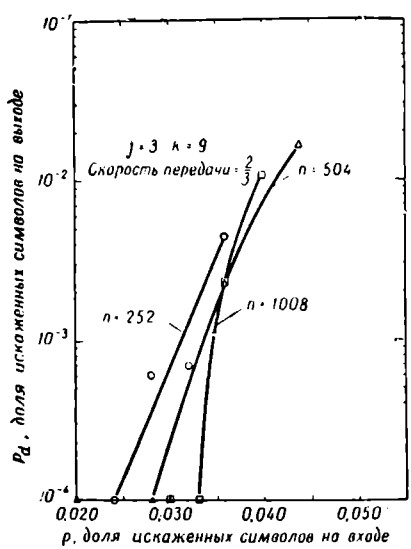
На рис. 6.5 сравниваются экспериментальные результаты, полученные при вероятностном декодировании кода с $n=504$, $j=3$ и $k=6$, с теоретически рассчи-



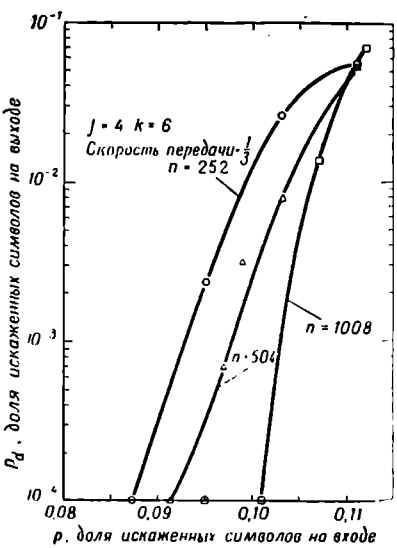
Р и с. 6.1. Экспериментальные результаты для ДСК.



Р и с. 6.2. Экспериментальные результаты для ДСК.



Р и с. 6.3. Экспериментальные результаты для ДСК.



Р и с. 6.4. Экспериментальные результаты для ДСК.

танной вероятностью ошибки декодирования того же кода по методу максимума правдоподобия. Для сравнения рассмотрен код Боуза — Чоудхури примерно с той же длиной блока и с той же скоростью. Для этого кода значение $P(e)$ получено в предположении использования одного из известных алгоритмов декоди-

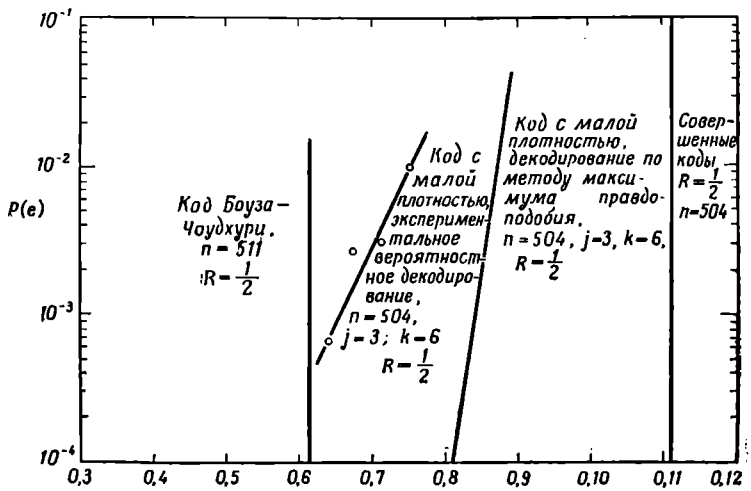


Рис. 6.5. Сравнение экспериментальных результатов вероятностного декодирования с теоретическими возможностями декодирования по методу максимума правдоподобия.

рования, например алгоритма Питерсона [12]. Эти алгоритмы позволяют исправлять число ошибок, не превосходящее половины минимального расстояния. Из графиков следует, что код Боуза — Чоудхури оказывается более эффективным при малых вероятностях перехода, а код с малой плотностью проверок на четность — при больших.

6.3. Канал с белым гауссовским шумом

В двух следующих разделах каждый из рассматриваемых каналов состоит из передатчика двоичной информации, физического канала и приемника,

работающего по методу максимума правдоподобия. Мы принимаем, что выходом этого приемника служит логарифмическое отношение правдоподобия

$$y = \ln \frac{\Pr [x = 0 \mid r(t)]}{\Pr [x = 1 \mid r(t)]},$$

где x — переданный символ, $r(t)$ — принятый сигнал, соответствующий этому символу, и y — выход приемника. Такой выход, конечно, можно было бы преобразовать в двоичный символ до того, как будет сделана попытка декодировать блок символов, однако при этом преобразовании была бы потеряна часть информации о переданной последовательности. Поскольку вероятностная схема естественным образом работает с логарифмическим отношением правдоподобий, естественно спросить, какой выигрыш в вероятности ошибки, мощности сигнала и скорости передачи можно получить при использовании декодером выхода приемника, вычисляющего отношение правдоподобия, вместо предварительно полученного двоичного решения. Оказывается, что для двух рассматриваемых ниже каналов этот выигрыш имеет первостепенное значение.

Допустим для канала с белым гауссовским шумом, что каждые T *сек* передается один из сигналов. Эти сигналы попадают на приемник должным образом ослабленными и с задержкой и представляются двумя функциями $x_0(t)$ и $x_1(t)$, отличными от нуля только в пределах от $t=0$ до $t=T$. Оба сигнала имеют равные энергии

$$E_c = \int_0^T x_0^2(t)\,dt = \int_0^T x_1^2(t)\,dt.$$

Пусть $n(t)$ — выборочное значение белого гауссовского шума со спектральной плотностью N_0 на единицу полосы; оно складывается с сигналом на приемнике. Легко показать тогда [8], что отношение правдоподобия, вычисленное идеальным приемником, равно

$$y = \frac{2}{N_0} \int_0^T [x_0(t) - x_1(t)]\, r t\, dt,$$

где $r(t)$ — принятый сигнал. Если $x=0$ есть переданный символ, то $r(t)=x_0(t)+n(t)$, и легко показать, что y нормально распределено с плотностью

$$\Pr(y|x=0)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(y-\sigma^2/2)^2}{2\sigma^2}\right], \qquad (6.1)$$

$$\sigma^2=\frac{4E_c(1-\rho)}{N_0}; \quad \rho=\frac{1}{E_c}\int_0^{T_0} x_0(t)\,x_1(t)\,dt. \qquad (6.2)$$

Аналогичным образом

$$P(y|x=1)=\frac{1}{\sqrt{2\pi\sigma}}-\frac{\left(y+\frac{\sigma^2}{2}\right)^2}{2\sigma^2}.$$

Эти вероятности иллюстрируются рис. 3.1, *в*.

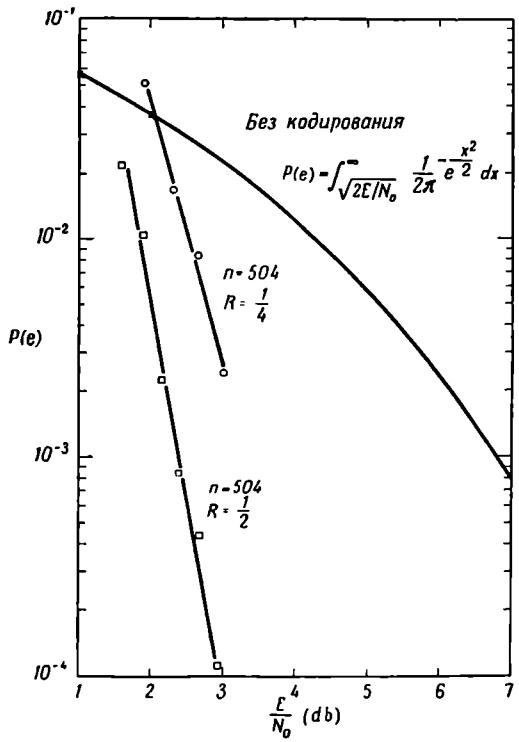
На вычислительной машине ИБМ-7090 был выполнен ряд экспериментов с кодами различных длин и скоростей; при этом выход канала выбирался датчиком псевдослучайных чисел в соответствии с плотностью распределения вероятностей (6.1), которая соответствует нулевому кодовому слову. Моделируемый декодер запоминал принятые слова, а затем пытался декодировать их по вероятностному методу декодирования. Результаты этих экспериментов для длины блока 504 и скоростей $^1/_4$ и $^1/_2$ приведены на рис. 6.6. Энергия сигнала, отложенная по оси абсцисс, есть энергия на информационный символ, поэтому

$$E=\frac{E_c}{R}. \qquad (6.3)$$

Предполагалось, что второй сигнал противоположен, т. е. в выражении (6.2) $\rho=-1$. Для случая ортогональных сигналов следует к каждому значению на оси абсцисс прибавить 3 *дб*.

Тот факт, что вероятность ошибки для кодов со скоростью $^1/_2$ оказывается меньше, чем для кодов со скоростью $^1/_4$, требует некоторых разъяснений. Рассмотрим две системы с равными мощностями сигнала и шума, равными длинами блока и числом информационных символов в секунду. Если в одной системе кодирование производится со скоростью $^1/_2$, а во

второй—со скоростью $^1/_4$, то время на блок при скорости $^1/_2$ в два раза больше времени при скорости $^1/_4$. Поэтому улучшение при скорости $^1/_2$ следует объяснить прежде всего большей длительностью периода.



Р и с. 6.6. Сравнение кодирования с малой плотностью проверок на четность с передачей без кодирования при белом гауссовском шуме.

С теоретической точки зрения больше оснований сравнивать случаи разных скоростей при заданном периоде кодовых ограничений, или заданной длительности кодовых ограничений, измеряемой числом информационных символов. Но так как стоимость использования декодера для кодов с малой плотностью проверок на четность определяется в первую очередь дли-

тельностью ограничений, измеряемой числом символов в канале, мы использовали в качестве основы для сравнений именно эту последнюю длительность

Рассмотрим теперь две системы — одну с кодированием со скоростью $^1/_2$, вторую вообще без кодирования, и пусть обе системы приводят к вероятности ошибки на символ, равной 10^{-3}, и обе передают одинаковое число информационных символов в секунду. Поскольку по оси абсцисс на рис. 6.6. отложена энергия на информационный символ, график показывает, что система с кодированием требует на (6,8—2,4) *дб*,

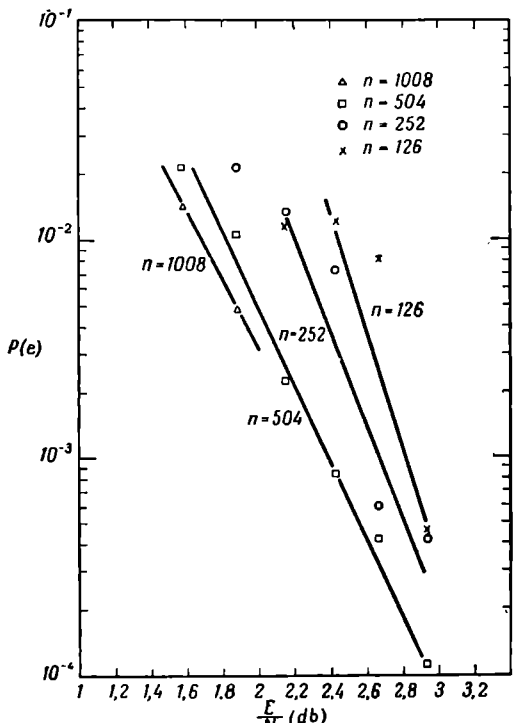


Рис. 6.7. Влияние длины блока на поведение кода со скоростью $^1/_2$ в канале с белым гауссовским шумом.

т. е. на 4,4 *дб*, меньшей мощности сигнала, чем система без кодирования. Скорость $^1/_4$ менее выгодна, так как увеличение возможности исправления ошибок все же не полностью компенсирует потери в энергии на символ сигнала в канале (рис. 6.7). Хотя и не было

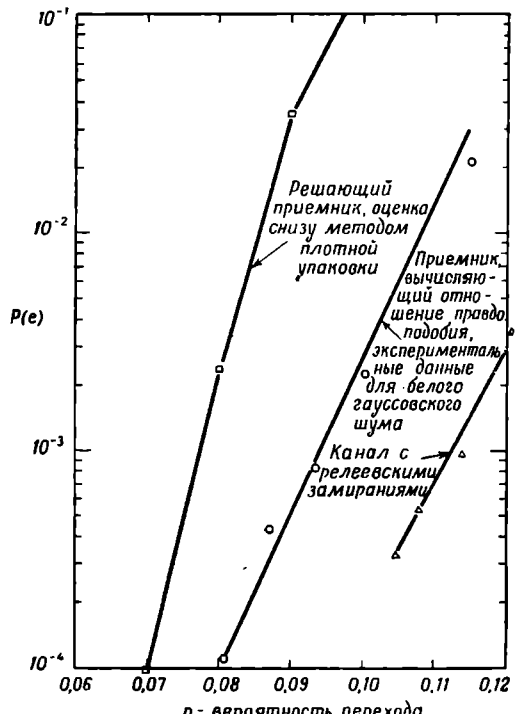


Рис. 6.8. Сравнение решающего приемника с приемником, вычисляющим отношения правдоподобия, $n = 504$, $R = {}^1/_4$.

получено экспериментальных результатов по использованию приемника рассматриваемого типа в случае кодов со скоростями $^1/_3$ и $^2/_3$, ввиду их малой эффективности при передаче по ДСК маловероятно, что они имеют какие-либо преимущества перед кодами со скоростью $^1/_2$.

Наконец, чтобы для декодирования проиллюстрировать преимущества приемника, вычисляющего отношение правдоподобия, перед решающим приемником, рассмотрим рис. 6.8. На нем сравниваются экспериментальные результаты применения кодов с малой плотностью проверок и вероятностного декодирования с оценкой снизу $P(e)$ для любого кода той же длины и той же скорости; оценка предполагает использование решающего приемника и декодирования по методу максимума правдоподобия. По оси абсцисс на рис. 6.7 отложена p — вероятность перехода, которая получится, если будет применен решающий приемник. Другими словами,

$$p = \int_{\sqrt{\frac{2E_c}{N_0}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Поучительным выводом из рис. 6.8 является важность использования приемника, вычисляющего отношение правдоподобия для увеличения корректирующей способности кода. Это в свою очередь показывает, что понятие «оптимальный код» не столь адекватно для техники связи, как это могло бы показаться из самого названия, и что простота и гибкость метода кодирования заслуживают большего внимания, чем свойство кода быть «оптимальным».

6.4. Канал с релеевскими замираниями

Допустим, что каждые T *сек* передается один из двух равновероятных, ортогональных, узкополосных сигналов равной энергии, и пусть $x_0(t)$ и $x_1(t)$ есть комплексное спектральное представление этих сигналов в положительной области частот. Допустим, что комплексное представление принятого сигнала имеет вид

$$r(t) = \alpha e^{j\beta} x_0(t) + n(t), \quad \text{если передано } x_0(t);$$
$$r(t) = \alpha e^{j\beta} x_1(t) + n(t), \quad \text{если передано } x_1(t),$$

где α имеет релеевское распределение, а β есть случайная фаза

$$\Pr(\alpha) = \alpha e^{-\frac{\alpha^2}{2}}, \quad \alpha \geqslant 0,$$

$$\Pr(\beta) = \frac{1}{2\pi}, \qquad 0 \leqslant \beta \leqslant 2\pi,$$

а $n(t)$ представляет собой белый гауссовский шум со спектральной плотностью N_0.

Можно показать, что, если до начала передачи ничего не известно относительно α и β, вся информация о том, какой из сигналов, $x_0(t)$ или $x_1(t)$, был передан, заключена в дискретизированных огибающих z_0 и z_1 на выходах фильтров, согласованных с $x_0(t)$ и $x_1(t)$:

$$z_0 = \left| \int_0^T x_0^*(t)\, r(t)\, dt \right| E_c^{-\frac{1}{2}},$$

$$z_1 = \left| \int_0^T x_1^*(t)\, r(t)\, dt \right| E_c^{-\frac{1}{2}}.$$

Пирс [13] показал, что z_0 и z_1 суть положительные случайные величины, распределенные по закону Релея с дисперсией $N_0 + E_c$ или N_0 в зависимости от того, что было передано, $x_0(t)$ или $x_1(t)$:

$$\Pr(z_i) = \frac{z_i}{N_0 + E_c} \exp\left[-\frac{z_i^2}{2(N_0 + E_c)} \right]$$

для $i = 0$ или 1, если передано $x_i(t)$; (6.4)

$$\Pr(z_i) = \frac{z_i}{N_0} \exp\left[-\frac{z_i^2}{2N_0} \right],$$

если передан сигнал, отличный от $x_i(t)$, (6.5)

где

$$E_c = \int_0^T x_0(t)\, x_0^*(t)\, dt. \tag{6.6}$$

Из независимости z_0 и z_1 и из выражений (6.4) и (6.5) непосредственно следует, что логарифмическое отношение правдоподобия на входе приемника имеет

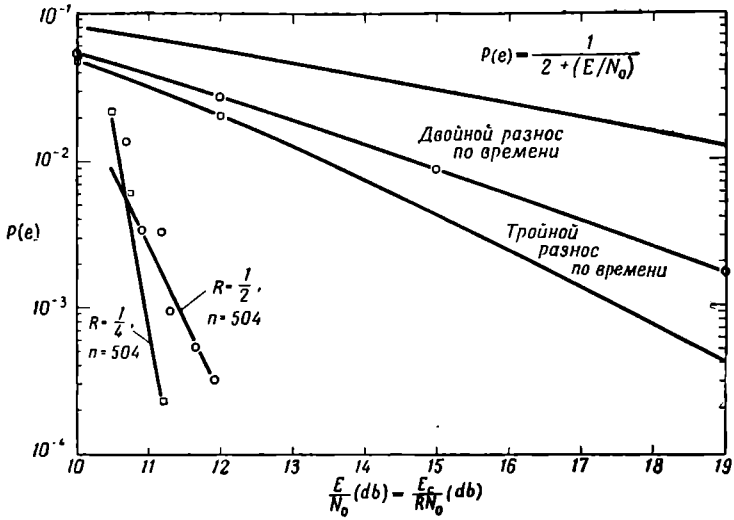


Рис. 6.9. Сравнение кодирования с малой плотностью проверок на четность с разносом по времени в канале с релеевскими замираниями.

следующий вид:

$$y=\ln\frac{p(x=0\mid z_0, z_1)}{p(x=1\mid z_0, z_1)}=(z_0^2-z_1^2)\left(\frac{1}{2N_0}-\frac{1}{2(N_0+E_c)}\right). \tag{6.7}$$

И наконец, из выражений (6.4), (6.5) и (6.7) получаем

$$\Pr(y\mid x=0)=\begin{cases}\dfrac{1+A}{A(2+A)}\,e^{-\frac{y}{A}}, & y\geqslant 0,\\ \dfrac{1+A}{A(2+A)}\,e^{-y\frac{(1+A)}{A}}, & y\leqslant 0,\end{cases} \tag{6.8}$$

где

$$A=\frac{E_c}{N_0}.$$

Канал с релеевскими замираниями моделировался на вычислительной машине при помощи датчика псевдослучайных чисел, задававшего выход y в соответствии с распределением (6.8). Последовательные значения y выбирались независимо, что, казалось бы, мало соответствует действительному положению вещей,

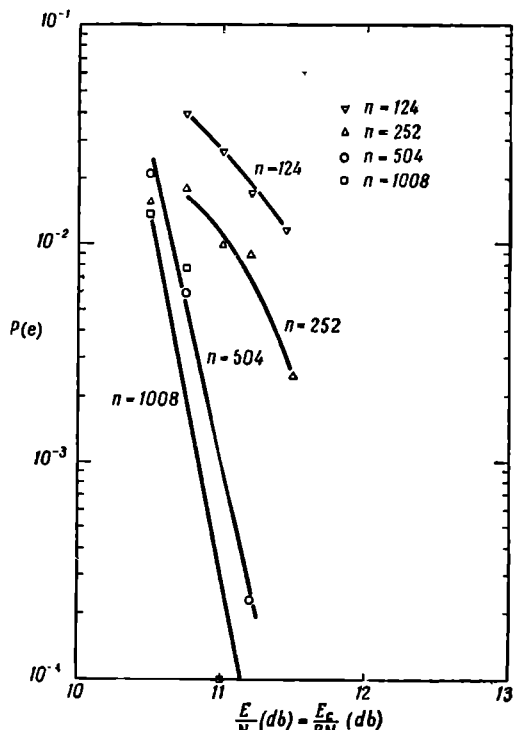


Рис. 6.10. Влияние длины блока на вероятность ошибки в канале с релеевскими замираниями при $R = 1/4$.

поскольку мы таким образом принимаем, что путь распространения сигнала не меняется за время T одного бода. Это допущение, однако, резонно, если скорость замираний сравнима с длительностью бода, и хоро-

шо отражает действительность, когда используется перемешивание символов последовательных блоков кода.

На рис. 6.9 приведены результаты моделирования. Эти графики еще убедительнее, чем графики рис. 6.6, показывают преимущества кодирования, что, конечно, связано с медленным убыванием частоты ошибок при увеличении мощности сигнала в канале с релеевскими замираниями. Рис. 6.9 показывает также, что код со скоростью $^1/_4$ несколько лучше кода со скоростью $^1/_2$, хотя мы и не имеем здесь совсем убедительных данных. К тому же код со скоростью $^1/_2$ содержит в два раза больше информационных символов на блок, чем код со скоростью $^1/_4$, так что при той же скорости передачи информации блок этого кода имеет в два раза большую длину. А это имеет преимущества, когда время замираний превышает продолжительность одного бода.

На рис. 6.10 показано влияние длины блока на вероятность ошибки для кода со скоростью $^1/_4$. Вероятность ошибки для кодов с меньшими длинами блока, по-видимому, убывает с ростом мощности сигнала значительно медленнее, чем для кодов с большой длиной; однако здесь нужны дальнейшие исследования.

И наконец, рис. 6.8 показывает преимущества приемника, вычисляющего отношение правдоподобия, по сравнению с решающим приемником для канала с релеевскими замираниями.

Гауссовские каналы и каналы с релеевскими замираниями столь различны по своим характеристикам, что можно выдвинуть гипотезу о том, что подобный выигрыш имеет место в большинстве симметричных каналов с двоичным входом (за очевидным исключением ДСК).

Приложение А

СВОЙСТВА ФУНКЦИИ $B(\lambda)$

В гл. 2 была получена следующая оценка функции распределения минимального расстояния (n, j, k)-ансамбля кодов

$$\left.\begin{aligned} &\Pr(D \leqslant n\delta) \leqslant \sum_{l=2}^{n} C(\lambda, n) \exp[-nB(\lambda)], \\ &\Pr(D \leqslant n\delta) \leqslant 1, \end{aligned}\right\} \quad (A.1)$$

где

$$\lambda = \frac{l}{n},$$

$$B(\lambda) = (j-1)H(\lambda) - \frac{j}{k}[\mu(s) + (k-1)\ln 2] + js\lambda, \quad (A.2)$$

$$C(\lambda, n) = [2\pi n\lambda(1-\lambda)]^{\frac{j-1}{2}} \exp \frac{j-1}{12n\lambda(1-\lambda)}, \quad (A.3)$$

$$\mu(s) + (k-1)\ln 2 = \ln \frac{1}{2}[(1+e^s)^k + (1-e^s)^k], \quad (A.4)$$

$$\frac{\mu'(s)}{k} = \lambda \text{ для оптимальной оценки.} \quad (A.5)$$

В настоящем приложении мы докажем три теоремы о неравенстве (A.1). В первой теореме исследуется поведение $B(\lambda)$, вторая дает оценку суммы в неравенстве (A.1) через ее первое и последнее слагаемое, а в третьей будет показано, что с ростом j и k оценка в неравенстве (A.1) сходится к функции распределения минимального расстояния ансамбля равновероятных кодов, оцениваемой в неравенстве (2.5)

Теорема А.1. *Пусть* $k>j\geqslant 3$, *и пусть* $B(\lambda)$ *определяется выражениями* (А.2), (А.4) *и* (А.5). *Тогда*

1. $\lim\limits_{\lambda\to 0} B(\lambda)=0$.
2. $\lim\limits_{\lambda\to 0}\dfrac{dB}{d\lambda}=\infty$.
3. $B(\lambda)$ *имеет всего один нуль в интервале* $0<\lambda<1/2$.
4. $B(\lambda)$ *не имеет локальных минимумов в области, где* $B(\lambda)>0$.

Доказательство. 1. Мы докажем, что $\lim\limits_{\lambda\to 0} B(\lambda)=0$, показав, что каждое из трех слагаемых правой части неравенства (А.2) стремится к нулю. Слагаемое $H(\lambda)$ задается выражением $-\lambda\ln\lambda-(1-\lambda)\ln(1-\lambda)$ и, как нетрудно видеть, стремится к 0. Дифференцируя обе части равенства (А.4), получаем

$$\lambda=\frac{\mu'(s)}{k}=\frac{e^s\left[(1+e^s)^{k-1}-(1-e^s)^{k-1}\right]}{(1+e^s)^k+(1-e^s)^k}, \qquad (A.6)$$

а отсюда $s\to-\infty$ при $\lambda\to 0$. Но из равенства (А.4) следует, что $\lim\limits_{s\to-\infty}\mu(s)+(k-1)\ln 2=0$, тогда

$$js\lambda=\frac{jse^s\left[(1+e^s)^{k-1}-(1-e^s)^{k-1}\right]}{(1+e^s)^k+(1-e^s)^k}.$$

Это выражение также стремится к 0 при $s\to-\infty$.

2. Из равенства (А.2) получаем

$$\frac{dB(\lambda)}{d\lambda}=\frac{\partial B(\lambda)}{\partial\lambda}+\frac{\partial B(\lambda)}{\partial s}\left(\frac{\partial\lambda}{\partial s}\right)^{-1}=(j-1)\ln\frac{1-\lambda}{\lambda}+js;$$

сделав подстановку

$$z=\frac{1-e^s}{1+e^s}, \quad s=\ln\frac{1-z}{1+z}, \qquad (A.7)$$

после некоторых преобразований равенства (А.6) получим

$$\lambda=\frac{1-z}{2}\,\frac{1-z^{k-1}}{1+z^k}. \qquad (A.8)$$

На рис. А.1 приведена зависимость s и λ от z.

$$\lim_{\lambda\to 0}\frac{dB}{d\lambda}=\lim_{z\to 1}(j-1)\ln\left(\frac{1+z}{1-z}\right)\left(\frac{1+z^{k-1}}{1-z^{k-1}}\right)+$$
$$+j\ln\frac{1-z}{1+z},$$
$$\lim_{\lambda\to 0}\frac{dB}{d\lambda}=\lim_{z\to 1}\ln\left(\frac{1-z}{1+z}\right)\left(\frac{1+z^{k-1}}{1-z^{k-1}}\right)^{j-1}, \qquad (A.9)$$
$$\lim_{\lambda\to 0}\frac{dB}{d\lambda}=\lim_{z\to 1}\ln\frac{(1+z^{k-1})^{j-1}}{(1-z^{k-1})^{j-2}(1+z)(1+z+\ldots+z^{k-2})},$$
$$\lim_{\lambda\to 0}\frac{dB}{d\lambda}=\infty \quad \text{при } j-2>0 \text{ или, другими словами, при } j\geqslant 3.$$

3. Прежде чем перейти к доказательству 3-й и 4-й частей теоремы, нужно показать, что $dB/d\lambda$ имеет

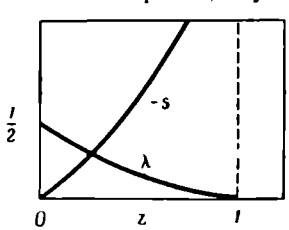


Рис. А.1. Зависимость s и λ от z.

всего один экстремум. Воспользовавшись равенством (А.9), получим производную $dB/d\lambda$ по z:

$$\frac{d}{dz}\left(\frac{dB}{d\lambda}\right)=\frac{2}{1-z^2}+\frac{2(j-1)(k-1)z^{k-2}}{1-z^{2(k-1)}}.$$

Положив ее равной нулю, получим

$$(j-1)(k-1)=\frac{(1-z^{2k-2})}{(1-z^2)z^{k-2}}=\frac{1+z^2+z^4+\ldots+z^{2k-4}}{z^{k-2}},$$

$$(j-1)(k-1)=1+\sum_{l=1}^{\frac{k-2}{2}}\left(z^{2l}+\frac{1}{z^{2l}}\right) \text{ при четных } k, \qquad (A.10)$$

$$(j-1)(k-1)=\sum_{l=1}^{\frac{k-1}{2}}\left(z^{2l-1}+\frac{1}{z^{2l-1}}\right) \text{ при нечетных } k. \qquad (A.11)$$

Функции в правых частях уравнений (А.10) и (А.11) убывают с ростом z при $0<z<1$. Поэтому каждое уравнение имеет самое большее одно решение в этом интервале. Таким образом, $dB/d\lambda$ имеет самое большее один экстремум и самое большее два нуля в интервале $0<\lambda<1/2$. Тогда, помимо $B(0)=0$, функция B обращается в нуль самое большее два раза. Но поскольку B положительна при λ, близких к нулю, то из того, что она обращается в нуль два раза в интервале $0<\lambda<1/2$, следует, что $B(1/2)>0$. Из равенства (А.4), однако, при $s=0$ в точке $\lambda=1/2$ получаем, что

$$B\left(\frac{1}{2}\right)=\left[(j-1)\ln 2-\frac{j}{k}(k-1)\ln 2-\left(1-\frac{j}{k}\right)\ln 2\right]<0.$$

Поэтому $B(\lambda)$ имеет в точности один нуль в интервале

$$0<\lambda<1/2.$$

4. Если бы $B(\lambda)$ имело минимум в области, где $B(\lambda)>0$, то отсюда следовало бы существование максимумов по обе стороны минимума, для того чтобы оказывались выполненными условия $B(0)=0$ и $B(1/2)<0$. Но это невозможно, так как $B(\lambda)$ имеет самое большее два экстремума. Ч.Т.Д.

Теорема А.2. *Функцию распределения минимального расстояния* (n, j, k)*-ансамбля можно оценить следующим образом*[1]:

$$\Pr(D\leqslant n\delta)\leqslant\frac{k-1}{2n^{j-2}}+o(n^{-j+2})+nC(\delta, n)\exp[-nB(\delta)]. \tag{А.12}$$

Доказательство. Из неравенства (2.18) получаем, что

$$\Pr(D<n\delta)\leqslant\sum_{l=2}^{n\delta}[C_n^l]^{-j+1}[N_1(l)]^j.$$

---

[1]) $o(n^{-j+2})$ означает функцию, убывающую с ростом n быстрее, чем n^{-j+2}, т. е. такую функцию $f(n)$, что $\lim_{n\to\infty} n^{j-2}f(n)=0$.

Слагаемое при $l=2$ можно оценить непосредственно. Вспомним, что $N_1(2)$ есть число последовательностей веса 2, удовлетворяющих первым n/k проверочным соотношениям любого фиксированного кода. Существует всего C_k^2 способов расположить две единицы в одном проверочном множестве; умножим число это на число проверочных множеств n/k и получим

$$N_1(2)=\frac{n}{k}C_k^2,$$

$$[C_n^2]^{-j+1}N_1(2)^j=\frac{n(k-1)^j}{2(n-1)^{j-1}}=\frac{(k-1)^j}{2n^{j-2}}+o(n^{-j+2}),$$

$$\Pr(D<n\delta)\leqslant\frac{(k-1)^j}{2n^{j-2}}+o(n^{-j+2})+$$

$$+\sum_{l=4}^{n\delta}C(\lambda,n)\exp[-nB(\lambda)], \quad \text{(A.13)}$$

где $C(\lambda, n)$ и $B(\lambda)$ задаются равенствами (А.2) и (А.3). Для того чтобы оценить слагаемые в неравенстве (А.13), для которых l мало, заметим, что из равенства (А.6) следует, что $s\to(1/2)\ln\lambda/(k-1)$ при $\lambda\to 0$. Подставим найденное значение вместо решения уравнения $\mu'(s)/k=\lambda$ в равенство (А.2); в этом случае $B(\lambda)$ оценивается снизу следующим образом:

$$B(\lambda)\geqslant(j-1)\left[\lambda\ln\frac{1}{\lambda}+(1-\lambda)\ln\frac{1}{1-\lambda}\right]+$$

$$+\frac{j}{k}\ln\sum_{l\text{ четные}}C_k^le^{sl}+\frac{j}{2}\lambda\ln\frac{\lambda}{k-1},$$

$$B(\lambda)\geqslant\left(\frac{j}{2}-1\right)\lambda\ln\frac{1}{\lambda}-\frac{j}{k}\ln\left[\frac{1}{1-C_k^2e^{2s}}\right]-\frac{j}{2}\lambda\ln(k-1). \quad \text{(A.14)}$$

Подставив l/n вместо λ и воспользовавшись некоторыми неравенствами, получим

$$\exp -nB(\lambda) \leqslant n^{-l\left(\frac{j}{2}-1\right)} l^{l\left(\frac{j}{2}-1\right)} (k-1)^{\frac{jl}{2}} \times \\ \times \exp\left(\frac{lj}{2}\right)\left(\frac{1}{1-\frac{kl}{2n}}\right). \quad \text{(A.15)}$$

Из равенства (А.3) имеем

$$C(\lambda, n) \leqslant (2\pi l)^{\frac{j-1}{2}} \exp\frac{j-1}{3n}. \quad \text{(A.16)}$$

Из неравенств (А.15) и (А.16) нетрудно видеть, что слагаемые с $l=4$ и $l=6$ в (А.13) убывают быстрее, чем n^{-j+2}. Из теоремы А.1 следует, что $B(\lambda)$ при $B(\delta)>0$ во всех слагаемых со значениями l от $l=8$ до $l=n\delta$ оценивается снизу величиной либо $B(8/n)$, либо $B(\delta)$ (если $B(\delta)<0$, правая часть неравенства (А.12) больше единицы и оценивается единицей). Таким образом, сумма от $l=8$ до $l=n\delta$ оценивается выражением

$$nC_{\max}\left\{\exp\left[-nB\left(\frac{8}{n}\right)\right]+\exp\left[-hB(\delta)\right]\right\}. \quad \text{(A.17)}$$

Первое слагаемое в выражении (А.17) имеет следующую зависимость от n:

$$n^{\left[1+\frac{\nu-1}{2}+8\left(-\frac{j}{2}+1\right)\right]} = o(n^{-j+2}) \quad \text{при} \quad j \geqslant 3.$$

Второе слагаемое в выражении (А.17) совпадает с последним слагаемым в правой части неравенства (А.12) в утверждении теоремы. Ч.Т.Д.

Теорема А.3. *Пусть δ_{jk} — отличное от нуля решение уравнения $B(\lambda)=0$ для (n, j, k)-ансамбля, и пусть фиксировано $R=1-j/k$. Пусть $\delta_0<1/2$ — решение уравнения $H(\delta_0)=(1-R)\ln 2$. Тогда* $\lim\limits_{k\to\infty} \delta_{jk}=\delta_0$.

Теорема 2.2 утверждает, что $\delta_0 n$ есть типичное минимальное расстояние в ансамбле равновероятных кодов с проверками на четность; таким образом, настоя-

щая теорема утверждает, что с ростом k типичное минимальное расстояние (n, j, k)-кода сходится к типичному минимальному расстоянию равновероятного ансамбля.

Доказательство. Перепишем выражение для $B(\lambda)$ в равенстве (А.2) следующим образом:

$$B(\lambda)=\left\{-H(\lambda)+\frac{j}{k}\ln 2\right\}+$$
$$+\left\{j[H(\lambda)+s\lambda]-\frac{j}{k}\ln[(1+e^s)^k+(1-e^s)^k]\right\}. \quad (A.18)$$

Покажем, что второе слагаемое в фигурных скобках в выражении (А.18) стремится к 0 с ростом k при $\lambda\neq 0$. Этого достаточно для доказательства теоремы, поскольку $j/k=1-R$, и потому первое слагаемое равно нулю только при $\lambda=\delta_0$:

$$H(\lambda)+s\lambda=\lambda\left[\ln\left(\frac{1-\lambda}{\lambda}\right)+s\right]-\ln(1-\lambda).$$

Сделав подстановку $z=(1-e)^s/(1+e)^s$ (см. равенства А.7 и А.8), получим

$$H(\lambda)+s\lambda=\frac{1-z}{2}\,\frac{1-z^{k-1}}{1+z^k}\,1\,\frac{1+z^{k-1}}{1-z^{k-1}}-$$
$$-\ln\left(\frac{1+z}{2}\right)\left(\frac{1+z^{k-1}}{1+z^k}\right) \quad (A.19)$$

и

$$\frac{1}{k}\ln[(1+e^s)^k+(1-e^s)^k]=\ln(1+e^s)+\frac{1}{k}\ln(1+z^k)=$$
$$=\ln\frac{2}{1+z}+\frac{1}{k}\ln(1+z)^k. \quad (A.20)$$

Объединяя равенства (А.19) и (А.20), получаем выражение для второго слагаемого в равенстве (А.18)

$$-j\left(\frac{1-z}{2}\right)\left(\frac{1-z^{k-1}}{1-z^k}\right)\ln\frac{1+z^{k-1}}{1-z^{k-1}}+$$
$$+j\ln\frac{1+z^{k-1}}{1+z^k}+\frac{j}{k}\ln(1-z^k).$$

С ростом k значения z^k и z^{k-1} стремятся к нулю при любом $z<1$ (где $\lambda>0$). Разлагая логарифмы, получаем

$$-j\left(\frac{1-z}{2}\right)\left(\frac{1-z^{k-1}}{1+z^k}\right)2z^{k-1}+yz^{k-1}(1-z)+$$

$$+\frac{jz^k}{k}+\text{слагаемые высших порядков.}$$

В этом выражении $j\to\infty$ линейно по k, но $z^{k-1}\to$ $\to 0$ экспоненциально. Поэтому второе слагаемое в равенстве (А.18) стремится к 0. Ч.Т.Д.

Приложение Б

МАТЕМАТИЧЕСКИЙ ВЫВОД РАЗЛИЧНЫХ РЕЗУЛЬТАТОВ ГЛ. 3.

Б.1. Оценки Чернова

Теорема 3.1. *Пусть* $Z=\sum_{i=1}^{n} z_i$ — *сумма* n *независимых случайных величин,* $P_i(z_i)$ — *плотность распределения* i-*й случайной величины, а* $g_i(s)=\int_{-\infty}^{\infty}\exp(sz_i)P_i(z_i)\,dz_i$ — *производящая функция моментов* i-*й случайной величины. Тогда для всех* $s\geqslant 0$, *таких, что* $g_i(s)$ *существует, справедливо неравенство*

$$\Pr(Z\geqslant nz_0)\leqslant \exp(-nsz_0)\prod_{i=1}^{n} g_i(s). \qquad (Б.1)$$

В случае дискретных z_i *справедливо то же утверждение, с тем исключением, что* $P_i(z_i)$ *суть вероятности, а интеграл, определяющий* $g_i(s)$, *заменяется суммой.*

Доказательство. Сумма Z есть также случайная величина с функцией распределения $F(Z)$ и производящей функцией моментов, равной

$$G(s)=\int_{-\infty}^{\infty}\exp(sZ)\,dF(Z)=\overline{\exp(sZ)}. \qquad (Б.2)$$

Из определения Z получаем

$$G(s)=\overline{\exp\left(s\sum_{i=1}^{n} z_i\right)}=\overline{\prod_{i=1}^{n}\exp sz_i}.$$

Так как случайные величины независимы,

$$G(s)=\prod_{i=1}^{n}\overline{\exp sz_i}=\prod_{i=1}^{n}g_i(s). \qquad (Б.3)$$

Из равенств (Б.2) и (Б.3) теперь получаем

$$\prod_{i=1}^{n}g_i(s)=\int_{-\infty}^{\infty}\exp(sZ)\,dF(Z)\geqslant\int_{nz_0}^{\infty}\exp(sZ)\,dF(Z). \qquad (Б.4)$$

для $s\geqslant 0$ и $Z\geqslant nz_0$, $sZ\geqslant snz_0$.

Таким образом,

$$\prod_{i=1}^{n}(g_i(\geqslant\exp(snz_0)\int_{nz_0}^{\infty}dF(Z)=\exp(snz_0)\Pr(Z\geqslant nz_0) \qquad (Б.5)$$

Перенеся сомножитель в другую часть неравенства, получаем утверждение теоремы — неравенство (Б.1). Теорема доказывается абсолютно так же и в случае дискретных z_i. Ч.Т.Д.

На первый взгляд может показаться, что из-за грубости оценки в неравенствах (Б.4) и (Б.5) получена довольно слабая оценка в неравенстве (Б.1). Это, однако, не так, если значение параметра выбрано правильно и если nz_0 больше среднего значения Z. Чтобы показать это, рассмотрим произведение $F(Z)e^{sZ}$. $F(Z)$ при больших n быстро растет в окрестности точки $\overline{Z}$. Можно, однако, рассматривать e^{sZ} как весовой множитель, который придает большой вес большим значениям Z. Поэтому произведение $F(Z)e^{sZ}$ будет иметь острый пик при некотором Z, большем $\overline{Z}$. Тонкость заключается в том, чтобы выбрать такое значение s, при котором этот пик имеет место в точке $Z=nz_0$. Аналитически это можно получить, взяв частную производную правой части неравенства (Б.1) по s и положив ее равной 0. При этом

$$nz_0=\sum_{i=1}^{n}\frac{1}{g_i(s)}\frac{\partial g_i(s)}{\partial s}. \qquad (Б.6)$$

Можно показать, что при таком выборе s оценка равенства (Б.6), называемая оценкой Чернова, имеет по крайней мере точное экспоненциальное поведение по n.

Теорема 3.2. *Пусть* z_i *и* w_i, $1 \leqslant i \leqslant n$, *суть пары случайных величин с плотностями распределения* $P_i(z_i, w_i)$. *Пусть производящая функция моментов пар имеет следующий вид:*

$$h_i(r, t) = \int\int \exp(rz_i + tw_i) P_i(z_i, w_i)\, dz_i\, dw_i. \qquad \text{(Б.7)}$$

Пусть пары случайных величин не зависят друг от друга. Определим Z и W следующим образом:

$$\begin{aligned} Z &= \sum_{i=1}^{n} z_i, \\ W &= \sum_{i=1}^{l} w_i, \quad l \leqslant n. \end{aligned} \qquad \text{(Б.8)}$$

Тогда для любых чисел z_0 и w_0

$$\Pr(Z \leqslant nz_0;\ W \leqslant nw_0) \leqslant$$

$$\leqslant \prod_{i=1}^{l} h_i(r, t) \prod_{i=l+1}^{n} h_i(r, 0) \exp[-n(rz_0 + tw_0)] \qquad \text{(Б.9)}$$

при любых $r \leqslant 0$, $t \leqslant 0$, таких, что $h_i(r, t)$ существует. Когда z и w дискретны, неравенство (Б.9) справедливо, если интегралы в (Б.7) заменить суммами, а плотности — вероятностями.

Доказательство. Пусть $F(Z, W)$ — функция распределения пары Z, W и производящая функция моментов Z, W имеет следующий вид:

$$H(r, t) = \begin{cases} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \exp(rZ + tW)\, dF(Z, W), \\ \overline{\exp(rZ + tW)}. \end{cases} \qquad \text{(Б.10)}$$

Используя равенства (Б. 8) и независимость выборочных значений, получаем

$$H(r, t)=\overline{\exp\left[\sum_{i=1}^{l}(rz_i+tw_i)+\sum_{i=l+1}^{n}rz_i\right]}=$$

$$=\left[\prod_{i=1}^{l}\overline{\exp(rz_i+tw_i)}\left(\prod_{i=l+1}^{n}\overline{\exp rz_i}\right)\right]=$$

$$=\prod_{i=1}^{l}h_i(r, t)\prod_{i=l+1}^{n}h_i(r, 0). \qquad \text{(Б. 11)}$$

Объединим равенства (Б. 10) и (Б. 11):

$$\prod_{i=1}^{l}h_i(r, t)\prod_{i=l+1}^{n}h(r, 0)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(rZ+tW)\,dF(Z, W)\geqslant$$

$$\geqslant\int_{Z=-\infty}^{nz_0}\int_{W=-\infty}^{nw_0}\exp(rZ+tW)\,dF(Z, W).$$

Для $r\leqslant 0$, $t\leqslant 0$, $Z\leqslant nz_0$ и $W\leqslant nw_0$ имеем:

$$\exp(rZ+tW)\geqslant\exp(rnz_0+tnw_0),$$

$$\prod_{i=1}^{l}h_i(r, t)\prod_{i=l+1}^{n}h_i(r, 0)\geqslant$$

$$\geqslant\exp(rnz_0+tnw_0)\Pr(Z\leqslant nz_0,\ W\leqslant nw_0).$$

Перенося сомножитель в другую часть неравенства, получим утверждение теоремы — неравенство (Б.9).

Ч. Т. Д.

Б.2. Оптимальное значение $f(y)$

Требуется найти выражение для $f(y)=f(-y)$, максимизирующее выражение

$$E(s, r, \lambda)=\frac{r}{s-r}\ln g(s)-$$

$$-\frac{s}{s-r}[B(\lambda)+\lambda\ln h(r)+(1-r)\ln g(r)], \qquad \text{(Б. 12)}$$

где

$$g(s)=\int_{-\infty}^{\infty} P_0(y)^{1-s} f(y)^s\, dy, \qquad \text{(Б. 13)}$$

$$h(r)=\int_{-\infty}^{\infty} P_0(y)^{\frac{1}{2}(1-r)} P_1(y)^{\frac{1}{2}(1-r)} f(y)^r\, dy. \qquad \text{(Б. 14)}$$

Если мы запишем $f(y)$ в виде $f(y)=f_0(y)+\varepsilon f_\varepsilon(y)$, то $f_0(y)$ даст максимум $E(s, r, \lambda)$ в том случае, когда $E(s, r, \lambda)$ имеет максимум по ε в точке $\varepsilon=0$ независимо от выбора $f_\varepsilon(y)$. Условие $f(y)=f(-y)$ будет выполнено автоматически, если мы перепишем интегралы в равенствах (Б.13) и (Б.14) как интегралы с пределами 0 и ∞. Тогда

$$g(s)=\int_0^{\infty} [P_0(y)^{1-s}+P_1(y)^{1-s}][f_0(y)+\varepsilon f_\varepsilon(y)]^s\, dy, \qquad \text{(Б. 15)}$$

$$h(r)=\int_0^{\infty} 2P_0(y)^{\frac{1}{2}(1-r)} P_1(y)^{\frac{1}{2}(1-r)} [f_0(y)+\varepsilon f_\varepsilon(y)]^r\, dy. \qquad \text{(Б. 16)}$$

Используя равенства (Б.15) и (Б.16), получаем

$$\frac{\partial E(s, r, \lambda)}{\partial \varepsilon}=\frac{rs}{(s-r)\,g(s)}\int_0^{\infty}(P_0^{1-s}+P_1^{1-s})(f_0+\varepsilon f_\varepsilon)^{s-1} f_\varepsilon\, dy-$$

$$-\frac{s\lambda r}{(s-r)\,h(r)}\int_0^{\infty} 2P_0^{\frac{1}{2}(1-r)} P_1^{\frac{1}{2}(1-r)} (f_0+\varepsilon f_\varepsilon)^{r-1} f_\varepsilon\, dy-$$

$$-\frac{s(1-\lambda)r}{(s-r)\,g(r)}\int_0^{\infty}(P_0^{1-r}+P_1^{1-r})(f_0+\varepsilon f_\varepsilon)^{r-1} f_\varepsilon\, dy. \qquad \text{(Б. 17)}$$

Записав правую часть равенства (Б.17) в виде одного интеграла, нетрудно убедиться в том, что он равен 0 при $\varepsilon=0$ независимо от выбора $f_\varepsilon(y)$ только в том случае, когда подинтегральная функция тожде-

ственно равна 0. Поэтому

$$\frac{1}{g(s)}\left(P_0^{1-s}+P_1^{1-s}\right)f_0^{s-1}-\frac{\lambda}{h(r)}\,2P_0^{\frac{1}{2}(1-r)}P_1^{\frac{1}{2}(1-r)}f_0^{r-1}-$$
$$-\frac{1-\lambda}{g(r)}\left(P_0^{1-r}+P_1^{1-r}\right)f_0^{r-1}=0, \qquad \text{(Б. 18)}$$

$$f_0(y)^{s-r}=\frac{\lambda h(r)^{-1}\left[2P_0(y)^{(1-r)/2}\,P_1(y)^{(1-r)/2}\right]}{g(s)^{-1}\left[P_0(y)^{1-s}+P_1(y)^{1-s}\right.}+$$
$$+\frac{(1-\lambda)\,g(r)^{-1}\left[P_0(y)^{1-r}+P_1(y)^{1-r}\right]}{g(s)^{-1}\left[P_0(y)^{1-s}+P_1(y)^{1-s}\right]}. \qquad \text{(Б. 19)}$$

Заметим, наконец, что, если значение $f_0(y)$, удовлетворяющее равенству (Б.19), умножить на произвольную константу, полученная величина по-прежнему будет удовлетворять равенству (Б.19) из-за компенсирующего изменения $g(r)$, $h(r)$ и $g(s)$. Отсюда после небольших преобразований получим равенство (3.40).

Покажем теперь, что это значение $f_0(y)$ дает локальный максимум $E(s, r, \lambda)$ по ε:

$$\frac{\partial^2 E(s,r,\lambda)}{\partial\varepsilon^2}=\frac{r}{(s-r)\,g(s)^2}\left\{g(s)\,\frac{\partial^2 g(s)}{\partial\varepsilon^2}-\left[\frac{\partial g(s)}{\partial\varepsilon}\right]^2\right\}-$$
$$-\frac{s\lambda}{(s-r)\,h(r)^2}\left\{h(r)\,\frac{\partial^2 h(r)}{\partial\varepsilon^2}-\left[\frac{\partial h(r)}{\partial\varepsilon}\right]^2\right\}-$$
$$-\frac{s(1-\lambda)}{(s-r)\,g(r)^2}\left\{g(r)\,\frac{\partial^2 g(r)}{\partial\varepsilon^2}-\left[\frac{\partial g(r)}{\partial\varepsilon}\right]^2\right\}. \qquad \text{(Б. 20)}$$

Рассмотрим первую скобку в равенстве (Б. 20):

$$g(s)\,\frac{\partial^2 g(s)}{\partial\varepsilon^2}-\left[\frac{\partial g(s)}{\partial\varepsilon}\right]^2\Bigg|_{\varepsilon=0}=$$
$$=s(s-1)\left[\int_0^\infty\left(P_0^{1-s}+P_1^{1-s}\right)f_0^s\,dy\right]\times$$
$$\times\left[\int_0^\infty\left(P_0^{1-s}+P_1^{1-s}\right)f_0^{s-2}f_\varepsilon^2\,dy\right]-$$
$$-s^2\left[\int_0^\infty\left(P_0^{1-s}+P_1^{1-s}\right)f_0^{s-1}f_\varepsilon\,dy\right]^2.$$

Но по неравенству Шварца

$$\left[\int_0^\infty (P_0^{1-s} + P_1^{1-s}) f_0^{s-1} f_\varepsilon \, dy\right]^2 \leqslant$$

$$\leqslant \left[\int_0^\infty (P_0^{1-s} + P_1^{1-s}) f_0^s \, dy\right] \left[\int_0^\infty (P_0^{1-s} + P_1^{1-s}) f_0^{s-2} f_\varepsilon^2 \, dy\right],$$

$$g(s) \frac{\partial^2 g(r)}{\partial \varepsilon^2} - \left[\frac{\partial g(s)}{\partial \varepsilon}\right]^2 \Big|_{\varepsilon=0} \geqslant$$

$$\geqslant -sg(s) \int_0^\infty (P_0^{1-s} + P_1^{1-s}) f_0^{s-2} f_\varepsilon^2 \, dy.$$

Таким же образом можно показать, что при $\varepsilon = 0$

$$h(r) \frac{\partial^2 h(r)}{\partial \varepsilon^2} - \left[\frac{\partial h(r)}{\partial \varepsilon}\right]^2 \geqslant$$

$$\geqslant -rh(r) 2 \int_0^\infty P_0^{\frac{1}{2}(1-r)} P_1^{\frac{1}{2}(1-r)} f_0^{r-2} f_\varepsilon^2 \, dy,$$

$$g(r) \frac{\partial^2 g(r)}{\partial \varepsilon^2} - \left[\frac{\partial g(r)}{\partial \varepsilon}\right]^2 \geqslant$$

$$\geqslant -rg(r) \int_0^\infty (P_0^{1-r} + P_1^{1-r}) f_0^{r-2} f_\varepsilon^2 \, dy.$$

Объединяя эти результаты и используя условия $s \geqslant 0$, $r \leqslant 0$, находим

$$\frac{\partial^2 E(s, r, \lambda)}{\partial \varepsilon^2}\Big|_{\varepsilon=0} \leqslant \frac{rs}{s-r} \int_0^\infty \left[\frac{1}{g(s)} (P_0^{1-s} + P_1^{1-s}) f_0^{s-2} f_\varepsilon^2 - \right.$$

$$- \frac{\lambda}{h(r)} P_0^{\frac{1}{2}(1-r)} P_1^{\frac{1}{2}(1-r)} f_0^{r-2} f_\varepsilon^2 -$$

$$\left. - \frac{1-\lambda}{g(r)} (P_0^{1-r} + P_1^{1-r}) f_0^{r-2} f_\varepsilon^2 \right] dy. \quad \text{(Б. 21)}$$

Сравним, наконец, подинтегральную функцию в выражении (Б.21) с выражением (Б.18); мы видим, что она тождественно равна 0. Таким образом,

$$\frac{\partial^2 E(s, r\lambda)}{\partial \varepsilon^2}\Big|_{\varepsilon=0} \leqslant 0,$$

и мы показали, что выражение (3.40) дает локальный максимум $E(s, r, \lambda)$ по $f(y)$.

Б.3. Исключение $f(y)$ из выражения для экспоненты

В этом разделе мы упростим выражение для $E(s, r, \lambda)$, задаваемое равенствами (Б.12), (Б.13) и (Б.14), исключив $f(y)$ из равенств (Б.13) и (Б.14) с помощью равенства (3.41), которое для удобства переписано здесь под номером (Б.22):

$$f(y) = \left[P_0(y)^{\frac{1}{2}(1-r)} + P_1(y)^{\frac{1}{2}(1-r)}\right]^{\frac{2}{s-r}} \times \\ \times \left[P_0(y)^{1-s} + P_1(y)^{1-s}\right]^{-\frac{1}{s-r}}. \quad \text{(Б. 22)}$$

Прежде всего в равенстве (Б. 12) прибавим и вычтем $s/(s-r)\ln[g(r)+h(r)]$. Это даст нам следующее:

$$E(s, r, \lambda) = \frac{r}{s-r}\ln g(s) - \frac{s}{s-r}\{B(\lambda) + \lambda \ln \alpha + \\ + (1-\lambda)\ln(1-\alpha) \ + \ln[g(r)+h(r)]\}, \quad \text{(Б. 23)}$$

$$\alpha = \ln \frac{h(r)}{g(r)+h(r)}. \quad \text{(Б. 24)}$$

Переписывая $g(s)$ и $g(r)+h(r)$ с помощью равенства (Б.22), получаем

$$g(s) = g(r) + h(r) = \\ = \int_0^\infty \left[P_0(y)^{1-s} + P_1(y)^{1-s}\right]^{-\frac{r}{s-r}} \times \\ \times \left[P_0(y)^{\frac{1}{2}(1-r)} + P_1(y)^{\frac{1}{2}(1-r)}\right]^{\frac{2s}{s-r}}. \quad \text{(Б. 25)}$$

Подставляя равенство (Б.25) в (Б.23) и записывая выражение для α, получаем равенства (3.43), (3.44) и (3.45).

Б.4. Упрощение выражения для экспоненты в случае ансамбля случайных кодов с проверками на четность

Из равенства (3.47) следует, что в этом случае $\beta(\alpha) = (1 - R)\ln 2$ не зависит от α. Поэтому выражение для $E(s, r)$ в равенстве (3.43) не зависит от α и может быть записано следующим образом:

$$E(s,\ r) = \frac{s}{s-r}(1 - R)\ln 2 -$$

$$- \ln \int \left(P_0^{1-s} + P_1^{1-s}\right)^{-\frac{r}{s-r}} \left(P_0^{\frac{1}{2}(1-r)} + P_1^{\frac{1}{2}(1-r)}\right)^{\frac{2s}{s-r}} dy. \quad (Б.26)$$

Если мы сделаем теперь подстановку $\rho = s/(s-r)$, $\sigma_1 = 1 - s$, $\sigma_2 = 1/2(1-r)$, то получим

$$E_1(\sigma_1, \rho) = \rho(1 - R)\ln 2 - \ln \int \left(P_0^{\sigma_1} + P_1^{\sigma_1}\right)^{1-\rho} \left(P_0^{\sigma_2} + P_1^{\sigma_2}\right)^{2\rho} dy, \quad (Б.27)$$

где

$$\sigma_2 = \frac{1 - \sigma_1(1-\rho)}{2\rho}.$$

Теперь найдем максимум $E_1(\sigma_1, \rho)$ по σ_1. Определим

$$z(\sigma_1,\ y) = [P_0(y)]^{\sigma_1} + [P_1(y)]^{\sigma_1},$$

$$E_1(\sigma_1,\ \rho) = \rho(1 - R)\ln 2 - \ln \int_0^\infty z(\sigma_1,\ y)^{1-\rho} z(\sigma_2,\ y)^{2\rho} dy,$$

$$\frac{\partial E_1(\sigma_1, \rho)}{\partial \sigma_1} = -\int_0^\infty z(\sigma_1, y)^{1-\rho} z(\sigma_2,\ y)^{2\rho} \times$$

$$\times \left[\frac{(1-\rho)\, z_1'(\sigma_1,\ y)}{z(\sigma_1,\ y)} - \frac{2\rho z_1'(\sigma_2,\ y)(1-\rho)}{z(\sigma_2,\ y)\, 2\rho}\right] dy \times$$

$$\times \left\{\int_0^\infty z(\sigma_1,\ y)^{1-\rho} z(\sigma_2,\ y)^{2\rho} dy\right\}^{-1}. \quad (Б.28)$$

Частная производная в равенстве (Б.28) взята при постоянном ρ, но при этом считается, что σ_2 изменяется в зависимости от σ_1 в соответствии с соотношением (Б.27). Из вида выражения в скобках в (Б.28) ясно, что $E_1(\sigma_1, \rho)$ имеет стационарную точку при $\sigma_1=\sigma_2$ или $1-s=\frac{1}{2}(1-r)$.

Чтобы показать, что $\sigma_1=\sigma_2$ действительно максимизирует $E_1(\sigma_1, \rho)$, достаточно показать, что выражение (Б.28) не отрицательно при $\sigma_1<\sigma_2$ и не положительно при $\sigma_1>\sigma_2$. Поскольку знак выражения (Б.28) определяется только членом в скобках, достаточно показать, что

$$\frac{\partial}{\partial\sigma_1}\left[\frac{-z_1'(\sigma_1, y)}{z(\sigma_1, y)}+\frac{z_1'(\sigma_2, y)}{z(\sigma_2, y)}\right]\leqslant 0 \qquad \text{(Б.29)}$$

или что

$$-\frac{z_1''(\sigma_1, y)\,z(\sigma_1, y)-[z_1'(\sigma_1, y)]^2}{[z(\sigma_1, y)]^2}+$$

$$+\frac{z_1''(\sigma_2, y)\,z(\sigma_2, y)-[z_1'(\sigma_1, y)]^2}{[z(\sigma_1, y)]^2}\left[-\frac{(1-\rho)}{2\rho}\right]\leqslant 0.$$

Перепишем первое слагаемое

$$-\frac{\left[P_0^{\sigma_1}(\ln P_0)^2+P_1^{\sigma_1}(\ln P_1)^2\right]\left[P_0^{\sigma_1}+P_1^{\sigma_1}\right]-\left[P_0^{\sigma_1}\ln P_0+P_1^{\sigma_1}\ln P_1\right]^2}{[z(\sigma_1, y)]^2}.$$

Но согласно неравенству Шварца первый член в знаменателе этого выражения не превышает второго, поэтому все первое слагаемое отрицательно. Из тех же соображений и второе слагаемое отрицательно, поэтому равенство $1-s=(1-r)/2$ дает максимум $E(s, r)$. Подставляя это значение в равенство (Б.26), получаем

$$E(s)=\frac{s}{1-s}(1-R)\ln 2-\ln\int_0^\infty\left[P_0^{1-s}+P_1^{1-s}\right]^{\frac{1}{1-s}}dy. \qquad \text{(Б.30)}$$

Б.5. Общий случай ДСК

Для того чтобы максимизировать выражение (3.61) по s и r и таким образом минимизировать оценку сверху $\bar{P}_e$ для ДСК, можно просто объединить равен-

ства (3.61), (3.62) и (3.63) и положить производные результата по s, r и λ равными 0. Это будет громоздко, и, кроме того, трудно показать, что найденная таким способом стационарная точка есть в самом деле максимум по s и r и минимум по λ. Вспомним, однако, что равенства (3.61) и (3.64) мы получили, исключая $f(y)$ из выражения в правой части неравенства (3.37).

В случае ДСК вид $f(y)$ не существен. Ввиду условий симметрии, задаваемых равенством (3.5), $f(+1) = f(-1)$, и, таким образом, $f(y)$ вполне определяется одним своим значением. Мы, кроме того, показали, что умножение $f(y)$ на константу не меняет величины E, поэтому для ДСК можно выбрать $f(y)$ равной 1. Теперь можно вернуться к неравенству (3.37) и минимизировать оценку непосредственно. Положив $p = P_0(-1)$, имеем

$$\overline{P}_e \leqslant \max_{\lambda} \min_{s, r, d} \{\exp n[\ln g(s) - sd] + \\ + nCn \exp n[B(\lambda) + \lambda \ln h(r) + (1-\lambda)\ln g(r) - rd]\}, \quad (Б.31)$$

$$\left.\begin{aligned} g(s) &= p^{1-s} + (1-p)^{1-s}, \\ h(r) &= 2p^{\frac{1-r}{2}}(1-p)^{\frac{1-r}{2}}. \end{aligned}\right\} \quad (Б.32)$$

Для минимизации выражения (Б.31) по s нужно просто минимизировать $[\ln g(s) - sd]$:

$$d = \frac{p^{1-s} \ln (1/p) + (1-p)^{1-s} \ln [1/(1-p)]}{p^{1-s} + (1-p)^{1-s}}, \quad (Б.33)$$

если d принадлежит области, в которой $0 \leqslant s \leqslant \infty$.

Чтобы показать, что равенство (Б.33) в самом деле минимизирует $P(e)$, мы можем проверить, что

$$\frac{\partial^2 [\ln g(s) - sd]}{ds^2} \geqslant 0. \quad (Б.34)$$

Это можно сделать либо прямым, но трудоемким дифференцированием, либо вспомнив, что вторая производная производящей функции семиинвариантов

$[\ln g(s)]$ всегда положительна [4]. Аналогичным образом, минимизируя по r, получаем

$$d = \frac{-\lambda}{2} \ln p(1-p) + \\ + \frac{(1-\lambda)\, p^{1-r} \ln (1/p) + (1-p)^{1-r} \ln [1/(1-p)]}{p^{1-r} + (1-p)^{1-r}}, \quad \text{(Б. 35)}$$

если d принадлежит области, в которой $-\infty < r \leqslant 0$. И наконец, минимум по d можно получить, приравняв две экспоненты

$$\ln \left(p^{1-s} + (1-p)^{1-s}\right) - sd = B(\lambda) + \lambda \ln 2 + \\ + \frac{\lambda (1-r)}{2} \ln p(1-p) + (1-\lambda) \ln \left[p^{1-r} + (1-p)^{1-r}\right] - rd. \quad \text{(Б. 36)}$$

Вообще говоря, можно воспользоваться соотношениями (Б.33), (Б.35) и (Б.36) для того, чтобы выразить s, r и d через λ, если существует решение при $0 \leqslant s < \infty$; $-\infty < r \leqslant 0$. Чтобы упростить эти соотношения, объединим прежде всего выражения (Б.33) и (Б.35) для того, чтобы исключить d:

$$-p_s \ln p - (1-p_s) \ln (1-p) = \\ = -\frac{\lambda}{2} \ln p(1-p) - (1-\lambda)\, p_r \ln p + (1-p_r) \ln (1-p), \quad \text{(Б. 37)}$$

где

$$\left.\begin{aligned} p_s &= \frac{p^{1-s}}{p^{1-s} + (1-p)^{1-s}}, \\ p_r &= \frac{p^{1-r}}{p^{1-r} + (1-p)^{1-r}}, \\ p_r &\leqslant p \leqslant p_s. \end{aligned}\right\} \quad \text{(Б. 38)}$$

Третье неравенство в (Б.38) необходимо для того, чтобы были удовлетворены неравенства $s \geqslant 0$ и $r \leqslant 0$.

Перегруппировав слагаемые равенства (Б.37), получим

$$\left(-p_s+\frac{\lambda}{2}+(1-\lambda)\,p_r\right)\ln p=$$
$$=\left[(1-p_s)-\frac{\lambda}{2}-(1-\lambda)(1-p_r)\right]\ln(1-p)=$$
$$=\left[-p_s+\frac{\lambda}{2}+(1-\lambda)\,p_r\right]\ln(1-p),$$

$$\boxed{p_s=\frac{\lambda}{2}+(1-\lambda)\,p_r} \qquad \text{(Б.39)}$$

Равенство (Б.36) упростится, если мы прибавим d к обеим его частям и подставим (Б.33) в правую часть (Б.36), а (Б.35) — в левую. После некоторых упрощений получим

$$\boxed{H(p_s)=B(\lambda)+\lambda\ln 2+(1-\lambda)H(p_r)} \qquad \text{(Б.40)}$$

Кроме того, поскольку мы приравняли экспоненты в правой части неравенства (Б.31), можно упростить выражение для $\bar{P}_e$. Поступая так же, как и при выводе равенства (Б.40), получаем

$$\boxed{\begin{aligned}&\bar{P}_e\leqslant\max_{\lambda}(1+nC_n)\times\\&\times\exp\left\{-n\left[-H(p_s)+p_s\ln\left(\frac{1}{p}\right)+(1-p_s)\ln\left(\frac{1}{1-p}\right)\right]\right\}\end{aligned}} \qquad \text{(Б.41)}$$

где p_s удовлетворяет (Б.39) и (Б.40) и $p_r\leqslant p\leqslant p_s$.

Из рис. 3.4 нетрудно увидеть, что оценка $\bar{P}_e$ в неравенстве (Б.41) убывает с ростом p_s; таким образом, минимизация по λ означает отыскание такого λ, при котором минимизируется значение p_s, удовлетворяющее выражениям (Б.39) и (Б.40). Более простое выражение для этого значения λ можно найти из неравенства (Б.31); λ выбирается так, чтобы максимизировать

$$\left[B(\lambda)+\lambda\ln\frac{h(r)}{g(r)}\right]=B(\lambda)+\frac{\lambda}{2}\ln 4p_r(1-p_r). \qquad \text{(Б.42)}$$

Приложение В

АНАЛИЗ ЧИСЛА НЕЗАВИСИМЫХ ИТЕРАЦИЙ ПРИ ДЕКОДИРОВАНИИ

В гл. 4 было получено неравенство (4.19) — асимптотическая оценка вероятности ошибки при использовании метода вероятностного декодирования. Оценка была получена как функция числа итераций при декодировании. В настоящем приложении будут выведены верхняя и нижняя оценки максимального числа итераций при декодировании, которые можно провести для (n, j, k)-кода до того, как нарушатся условия независимости теоремы 4.1. Покажем сначала, что в любом (n, j, k)-коде m оценивается сверху следующим образом:

$$m < \frac{\log n}{\log (k-1)(j-1)}. \qquad \text{(B. 1)}$$

Затем, и это важнее, мы приведем конструктивный прием, всегда позволяющий найти (n, j, k)-код, удовлетворяющий неравенствам

$$m+1 > \frac{\log n + \log \frac{kj-k-j}{2k}}{2 \log (k-1)(j-n)} \geqslant m. \qquad \text{(B. 2)}$$

Заметим, что при больших n значение m, удовлетворяющее неравенству (B.2), равно примерно половине значения m, удовлетворяющего неравенству (B.1).

Теорема B.1. *Пусть m есть максимальное число независимых итераций при декодировании для кодов с длиной блока n, с проверочными множествами по k символов и j проверочными множествами на символ. Тогда*

$$m < \frac{\log n}{\log (k-1)(j-1)}.$$

Доказательство. Рассмотрим m-ярусное дерево проверочных множеств некоторого символа в (n, j, k)-коде. Для того чтобы можно было сделать m независимых итераций, всем узлам дерева должны соответствовать разные символы кода. Таким образом, число узлов в m-ярусном дереве должно быть не больше длины блока n. Первый ярус содержит по $k-1$ узлов на каждую из j ветвей, выходящих из корня. Таким образом, первый ярус содержит $j(k-1)$ символов. Каждый из этих символов дает $(j-1)(k-1)$ узлов во втором ярусе, так как всего j—1 ветвей выходят из каждого символа во втором ярусе. Таким образом, во втором ярусе всего $j(j-1)(k-1)^2$ символов. Аналогично в i-м ярусе содержится $j(j-1)^{i-1}(k-i)^i$ символов. Поэтому

$$1+j(k-1)+j(j-1)(k-1)^2+\ldots$$

$$\ldots+j(j-1)^{i-1}(k-1)^i+\ldots$$

$$\ldots+j(j-1)^{m-1}(k-1)^m\leqslant n. \quad \text{(В.3)}$$

Выражение в левой части неравенства (В.3) оценивается снизу последним слагаемым, а оно в свою очередь оценивается снизу выражением $(j-1)^m(k-1)^m$, отсюда

$$(j-1)^m(k-1)^m<n. \quad \text{(В.4)}$$

Взяв логарифмы обеих частей неравенства (В.4), получим неравенство (В.1), что и доказывает теорему. Ч. Т. Д.

Можно, кроме того, получить точную сумму в неравенстве (В.3):

$$1+j(j-1)^{m-1}(k-1)^m\left[\frac{1-(j-1)^{-m}(k-1)^{-m}}{1-(j-1)^{-1}(k-1)^{-1}}\right]\leqslant n. \quad \text{(В.5)}$$

Однако неравенство (В.5) громоздко и потому менее удобно.

Прежде чем перейти к методу построения кода с параметрами, удовлетворяющими неравенствам (В.2), установим связь между m и относительным расположением единиц в проверочной матрице. Назовем *за-*

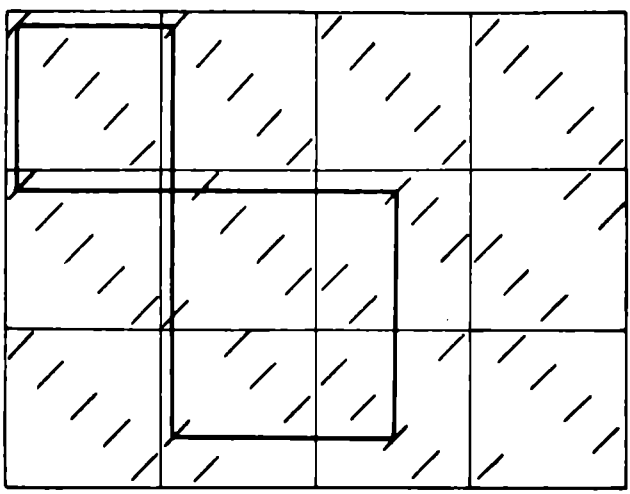


Рис. В.1. Пример замкнутого пути длины 6, штрихи — единицы, свободное место — нули.

мкнутым путем в проверочной матрице последовательность чередующихся горизонтальных и вертикальных отрезков со следующими свойствами. Во-первых, последний отрезок последовательности оканчивается в начале первого отрезка, и, во-вторых, вершина каждого угла находится в точке, где проверочная матрица содержит 1. Вершина угла есть, по определению, общая точка двух смежных отрезков последовательности; это определение включает и общую точку между последним и первым отрезками (см. рис. В.1). По определению, длина замкнутого пути есть число отрезков в последовательности. Например, замкнутый путь на рис. В.1 имеет длину 6. Заметим, что горизонтальные

и вертикальные отрезки могут пересекать другие отрезки и проходить через другие единицы в матрице и тем не менее подсчитываются один раз. Примем, что последовательность отрезков, образующая замкнутый путь, может начинаться с любых отрезков и может отсчитываться в любом направлении.

Лемма В.1. *Если существует один или большее число замкнутых путей длины L в проверочной матрице и не существует замкнутых путей длины, большей L, то число независимых итераций декодирования m удовлетворяет неравенствам*

$$m \leqslant \frac{L}{4} \leqslant m+1. \qquad \text{(В.6)}$$

Доказательство первой половины неравенства. Рассмотрим фиксированный замкнутый путь длины L. Тогда существует $L/2$ вертикальных отрезков, каждый из которых соответствует символу в коде. Перенумеруем эти символы в порядке их появления вдоль замкнутого пути: $a_1, a_2, \ldots, a_{L/2}$. Для замкнутого пути, показанного на рис. В.1, мы имели бы $a_1=1$, $a_2=6$, $a_3=13$. Рассмотрим дерево проверочных множеств, связанное с символом $a_{L/2}$ (см. рис. В.2), и рассмотрим два пути в этом дереве,

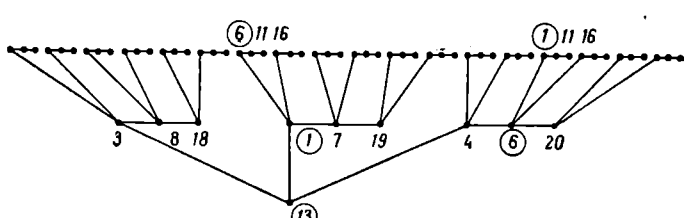


Рис. В.2. Замкнутый путь рис. В.1 в дереве проверочных множеств.

образованные $a_{L/2}, a_{(L/2)-1}, \ldots, a_{L/4}$ (или $a_{(L/4)+1/2}$) и $a_{L/2}, a_1, a_2, \ldots, a_{L/4}$ (или $a_{(L/4)+1/2}$). Заметим, что $a_{L/2}$ появляется в ярусе 0, a_1 и $a_{(L/2)-1}$ — в ярусе 1 и вообще a_i и $a_{(L/2)-i}$ — в ярусе i. Если $L/4$ — целое,

$a_{L/4}$ должно появиться в $L/4$-м ярусе дважды, и поэтому $m < L/4$. С другой стороны, если $L/4 + 1/2$ целое, $a_{(L/4)+1/2}$ появится один раз в $[L/4 - 1/2]$-м ярусе и один раз в $[L/4 + 1/2]$-м ярусе. В этом случае $m \leqslant L/4 - 1/2 < L/4$, а это и заканчивает доказательство того, что $m < L/4$.

Доказательство второй половины неравенства. Если возможны только m независимых итераций при декодировании, то в коде существует некоторый символ, скажем d, такой, что дерево проверочных множеств содержит некоторый символ, скажем a_0, в $(m+1)$-м ярусе, и этот символ появляется еще раз в $(m+1)$-м ярусе или на более низких ярусах. Пусть теперь a_1 и b_1 — символы, стоящие непосредственно под двумя символами a_0 в дереве проверочных множеств, символы a_2 и b_2 суть непосредственно следующие за ними и т. д. вплоть до символа d. Число символов в совокупности a_0, d, a_1, ..., b_1, ... не превышает $2(m+1)$. Нарисуем, наконец, замкнутый путь в проверочной матрице, начинающийся с a_0, и проверочное множество, содержащее a_1 и a_2, и дальше до символа d, а затем обратно через символы b. Такой замкнутый путь содержит в два раза большее число отрезков, чем символов; поэтому $L \leqslant 4(m+1)$, что и доказывает лемму. Ч.Т.Д.

Теперь опишем метод построения проверочной матрицы, не содержащей замкнутых путей длины $L=4m$ или меньшей. Затем последует доказательство того, что такое построение можно провести во всех тех случаях, когда удовлетворяется неравенство (В.2). Рассмотрим $(nj/k \times n)$-матрицу, такую, как приведенная на рис. В.3. Матрица разбита на jk квадратных подматриц с n/k строками и столбцами в каждом. Первая строка из подматриц и первый столбец из подматриц состоят целиком из единичных матриц. Другие подматрицы содержат символ U во всех позициях на главной диагонали и символ A во всех остальных позициях. Наша задача состоит в том, чтобы заменить все подматрицы, содержащие символы A

и U, перестановками $(n/k \times n/k)$-единичной матрицы так, чтобы не получить при этом замкнутые пути длины $4m$ или меньшей. Символ A использован для обозначения позиций, допустимых в том смысле, что мы можем поместить в них 1, не получая при этом замкнутые пути длины $4m$ или меньшей. Символ U указывает недопустимые позиции—это те позиции, помещение 1 в которые привело бы к образованию замкнутого пути длины $4m$ или меньшей. Заметим, что

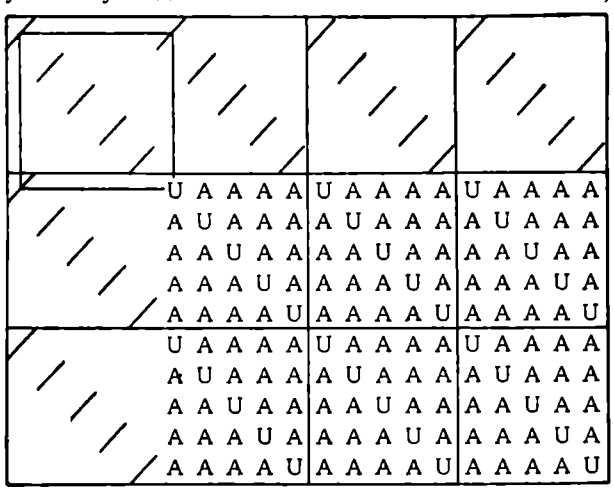


Рис. В.3. Первый шаг метода построения матрицы.

даже в случае $m=1$ главные диагонали содержат символы U из-за замкнутых путей длины 4, таких, какой показан на рис. В.3.

Возьмем теперь подматрицу, содержащую символы A и U; выберем в первой строке позицию с символом A и заменим его на 1. Кроме того, припишем 0 перед каждым символом подматрицы, находящимся в той же строке и том же столбце, что и 1. В тех позициях матрицы, в которых 1 не может быть помещена без того, чтобы при этом не образовался замкнутый путь длины $4m$ или меньшей, заменим A символом U. Это построение дает матрицу, такую, как приведенная на рис. В.4.

/					/					/					/				
	/					/					/					/			
		/					/					/					/		
			/					/					/					/	
				/					/					/					/
/					0U	/	0A	0A	0A	U	U	A	A	A	U	U	A	A	A
	/				A	0U	A	A	A	A	U	A	A	A	A	U	A	A	A
		/			A	0A	U	A	A	A	A	U	A	A	A	A	U	A	A
			/		A	0A	A	U	A	A	A	A	U	A	A	A	A	U	A
				/	A	0A	A	A	U	A	A	A	A	U	A	A	A	A	U
/					U	U	A	A	A	U	A	A	A	A	U	A	A	A	A
	/				A	U	A	A	A	A	U	A	A	A	A	U	A	A	A
		/			A	A	U	A	A	A	A	U	A	A	A	A	U	A	A
			/		A	A	A	U	A	A	A	A	U	A	A	A	A	U	A
				/	A	A	A	A	U	A	A	A	A	U	A	A	A	A	U

Рис. В.4. Второй шаг метода построения матрицы.

Продолжая процесс над строкой 2 подматрицы, заменим некоторые позиции, содержащие символ A (но не $0A$), на 1, заполним строку и столбец нулями и заменим символы A на U там, где это необходимо. Продолжим построение до тех пор, пока каждая строка подматрицы не будет содержать 1, и проделаем то же самое для каждой подматрицы. Если на некотором

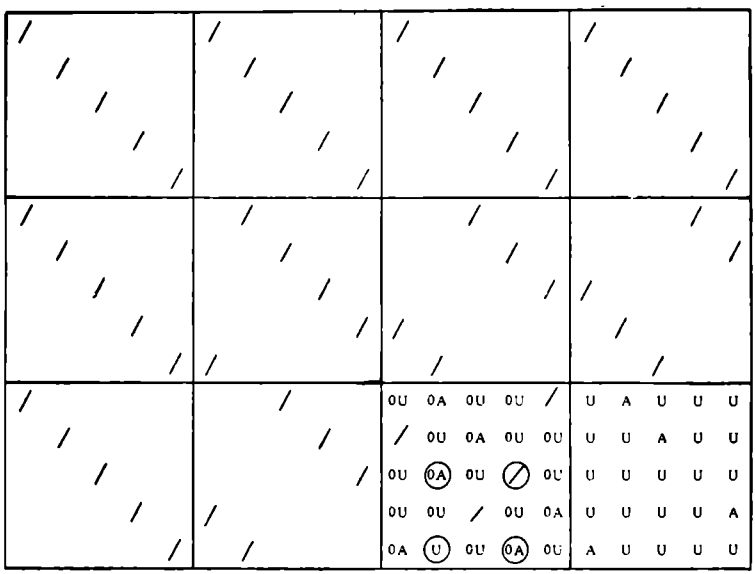


Рис. В.5. Пример „аварийного" приема при построении матрицы.

этапе этого процесса попадется строка, скажем l-я, не содержащая символа A без приписанных к нему нулей, выполним следующий «аварийный» прием. Пусть c_l — столбец, в котором строка l содержит символ U. Обозначим это следующим образом: $P(l, c_l) = U$, где $P(i, j)$ есть, по определению, символ, стоящий на пересечении i-й строки и j-го столбца подматрицы. Для каждого $i<l$ определим c_i как столбец, для которого $P(i, c_i)=1$. Найдем теперь такое i, для которого $P(i, c_l)=0A$ и $P(l, c_i)=0A$ (см. символы, за-

ключенные в кружок, на рис. В.5). Для этого i заменим $P(i, c_l)$ на 1, $P(l, c_i)$ на 1, $P(i, c_i)$ на $0A$ и изменим символы A, U на $0A$ во всей матрице в соответствии с новой совокупностью единиц.

Для того чтобы показать, что этот «аварийный» прием выполним, необходимо прежде всего показать, что, добившись одновременно того, что $P(i, c_l)=1$ и $P(l, c_i)=1$, мы не получим замкнутых путей длины $4m$ или меньшей. Кроме того, необходимо показать, что если неравенство (В.2) удовлетворяется, то всегда существует такое i, что $P(i, c_l)=0A$ и $P(l, c_i)=0A$.

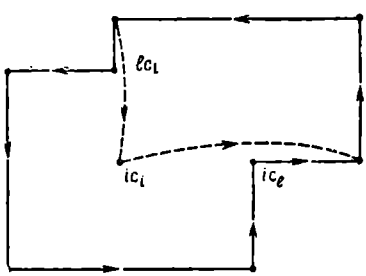


Рис. В.6. Случай 1: замкнутый путь через lc_l и lc_l.

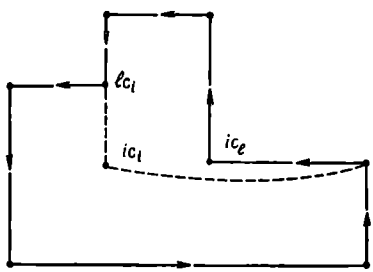


Рис. В.7. Случай 2: замкнутый путь через lc_l и ic_l.

Первое мы докажем от противного. Допустим, что, положив $P(i, c_l)=1$, $P(l, c_i)=1$ и $P(i, c_i)=0$, мы образовали замкнутый путь длины $4m$ или меньшей. Этот путь должен содержать обе точки (i, c_l) и (l, c_i) как вершины углов, поскольку стоящие ранее в этих позициях символы $0A$ указывали на то, что не существовало замкнутых путей длины $4m$ или меньшей и проходящих через каждую из этих точек в отдельности. Проследим этот замкнутый путь начиная с точки (l, c_i) вдоль по горизонтальной линии. Следует рассмотреть два случая: 1) путь проходит к (i, c_l) вдоль горизонтальной линии, как на рис. В.6, и 2) путь проходит к (i, c_l) вдоль вертикальной линии, как на рис. В.7.

В первом случае положим $P(i, c_l)=0$, $P(i, c_i)=1$ и оборвем горизонтальную линию, идущую к (i, c_l) в точке (i, c_i), как на рис. В.6. Замкнем путь,

двигаясь вертикально к (l, c_i). Этот путь имеет длину, меньшую чем $4m$, поскольку он короче первоначального пути. Однако это противоречит допущению о том, что (l, c_i) была допустимой точкой, когда $P(i, c_i)$ было равно 1.

Во втором случае положим $P(i, c_l)=0$, $P(l, c_i)=0$ и $P(i, c_i)=1$. Теперь оборвем в точке (i, c_i) вертикальную линию, раньше кончавшуюся в точке (l, c_i) (см. рис. В.7), а горизонтальную линию, раньше начинавшуюся в (i, c_l), начнем в точке (i, c_i) (см. рис. В.7). Мы получим таким образом замкнутый путь длины, меньшей чем $4m$, не включающий ни (i, c_l), ни (l, c_i). Это тоже противоречие, поскольку ни одна единица не была размещена в матрице так, чтобы образовался замкнутый путь длины $4m$ или меньшей. Этим и заканчивается доказательство того, что $P(l, c_i)$ и $P(i, c_l)$ можно одновременно положить равными 1, когда обе они обозначены $0A$.

Чтобы закончить доказательство, мы должны показать, что в том случае, когда удовлетворяется неравенство (В.2), всегда можно в «аварийной» ситуации найти такое i, что $P(l, c_i)=0A$ и $P(i, c_l)=0A$. Сначала покажем, что из неравенства (В.2) следует существование большего чем $n/2k$ числа значений i, при которых $P(l, c_i)=0A$. Затем покажем, что больше чем $n/2k$ элементов столбца c_l подматрицы суть символы $0A$. Эти два соотношения завершают доказательство, поскольку в том случае, когда $P(i, c_l)\neq 0A$ для всех i, для которых $P(l, c_i)=0A$, столбец c_l подматрицы содержит больше чем $n/2k$ элементов, отличных от $0A$, и больше чем $n/2k$ элементов $0A$. Но это невозможно, поскольку столбец l подматрицы содержит всего n/k элементов. Таким образом, должно существовать такое i, что $P(i, c_l)=0A$ и $P(l, c_i)=0A$.

Оценим теперь число точек строки l, которые могут быть обозначены символом U. Если некоторая точка l-й строки подматрицы недопустима, то помещенная в этой точке 1 приведет к образованию замкнутого пути длины $4m$ или меньшей. Рассмотрим случай, когда первый отрезок этого замкнутого пути есть горизонтальный отрезок вдоль строки l, начи-

нающийся в недопустимой точке. Тогда последним отрезком будет вертикальный отрезок, оканчивающийся в недопустимой точке. Мы хотим узнать сначала, сколько может быть замкнутых путей длины 4, начинающихся в недопустимой точке строки l подматрицы. Существует самое большее k—1 точек, в которых может заканчиваться первый горизонтальный отрезок, а именно это те позиции l-й строки полной матрицы, в которых уже были размещены единицы. Любой вертикальный отрезок, выходящий из одной из этих k—1 точек, может заканчиваться самое большее в i—1 точках, а именно в оставшихся позициях столбца, в которых размещены единицы. Вспомним, что при такой конструкции мы никогда не размещали больше чем k единиц в строке и j единиц в столбце. И наконец, если мы хотим построить замкнутый путь длины 4, следующий горизонтальный отрезок, выходящий из одной из этих j—1 точек, должен заканчиваться в столбце рассматриваемой подматрицы, но существует самое большее одна единица в любом из таких столбцов, в которой может заканчиваться рассматриваемый горизонтальный отрезок. Таким образом, существует самое большее $(k-1)(j-1)$ различных замкнутых путей длины 4, для которых некоторая недопустимая точка строки l заданной подматрицы может служить вершиной угла. Следовательно, из-за этих замкнутых путей длины 4 самое большее $(k-1)(j-1)$ точек строки l заданной подматрицы оказываются недопустимыми. Те же аргументы можно использовать при рассмотрении замкнутых путей длины 6. Здесь для любого из $(k-1)(j-1)$ путей длины 2 существует самое большее k—1 таких точек, в которых может заканчиваться третий отрезок, а для каждой из них существует самое большее j—1 точек, в которых может заканчиваться четвертый отрезок. Пятый отрезок теперь вполне определен, поскольку он должен заканчиваться в столбце заданной подматрицы. Отсюда самое большее $(k-1)^2(j-1)^2$ точек строки в заданной подматрице недопустимы из-за замкнутых путей длины 6. Аналогично замкнутые пути длины $2i$ приводят к тому, что самое большее

$(k-1)^{i-1}(j-1)^{i-1}$ точек оказываются недопустимыми. Следовательно, общее число недопустимых точек строки l заданной подматрицы N_u оценивается следующим образом:

$$N_u \leqslant \sum_{i=2}^{2m} (k-1)^{i-1}(j-1)^{i-1} =$$

$$= (k-1)^{2m-1}(j-1)^{2m-1}\left\{\frac{1-[(k-1)(j-1)]^{-(2m-1)}}{1-[(k-1)(j-1)]^{-1}}\right\} <$$

$$< \frac{[(k-1)(j-1)]^{2m-1}}{1-[(k-1)(j-1)]^{-1}} = \frac{[(k-1)(j-1)]^{2m}}{kj-k-j}. \quad \text{(B.7)}$$

Таким образом, $N_u < n/2k$, если

$$\frac{[(k-1)(j-1)]^{2m}}{kj-k-j} \leqslant \frac{n}{2k}$$

или

$$m \leqslant \frac{\log n + \log \dfrac{kj-k-j}{2k}}{2\log(k-1)(j-1)}.$$

Поскольку все элементы строки l заданной подматрицы суть либо $0A$, либо U, из неравенства (B.2) следует, что число элементов строки l, обозначенных $0A$, больше чем $n/2k$.

И наконец, мы должны показать, что число элементов столбца c_l подматрицы, обозначенных $0A$, больше чем $n/2k$. Доводы здесь те же, за исключением того, что вместо построения замкнутых путей, начинающихся в недопустимой точке с горизонтального отрезка, мы строим пути, начиная с вертикального отрезка. Неравенство (B.7) по-прежнему дает оценку числа недопустимых точек, а в силу неравенства (B.2) больше чем $n/2k$ точек обозначены $0A$, поскольку все элементы столбца c_l суть либо U, либо $0A$. Таким образом, мы указали конструктивный метод построения кода с m независимыми итерациями при декодировании, где m удовлетворяет неравенству (B.2).

ЛИТЕРАТУРА

1. Bloom F. J., Chang S. S. L., Harris B., Hauptschein A., Morgan K. C., Improvement of Binary Transmission by Null-Zone Reception, *Proc. IRE,* **45** (1957), 963—975.

2. Bose R. C., Ray-Chaudhuri D. K., On a Class of Error-Correcting Binary Group Codes, *Inf. and Control.,* **3** (1960), 68—79. (Русский перевод: Боуз Р. К., Рой-Чоудхури Д. К., Об одном классе двоичных групповых кодов с исправлением ошибок, Кибернетический сборник, вып. 2, ИЛ, М., 1961, 83—94.)

3. Elias P., Coding for Two Noisy Channels, Information Theory, Cherry C. (ed.), Third London Symposium, September, 1955, Butterworth's Scientific Publications, London, England. (Русский перевод: Элайес П., Кодирование для двух каналов с шумами, в сб. Теория передачи сообщений, ИЛ, М., 1957, 114—138.)

4. Fano R. M., Transmission of Information, M. I. T. Press and Wiley, New York, 1961. (Русский перевод: Фано Р., Передача информации. Статистическая теория связи, «Мир», М., 1965.)

5. Fano R. M., A Heuristic Discussion of Probabalistic Decoding, *IEEE Trans,* **IT-9**, 62—71 (1963).

6. Gilbert E. N., A Comparison of Signaling Alphabets, *Bell System Tech. J.,* **31** (1952), 504—522.

7. Гнеденко Б. В., Колмогоров А. Н., Предельные распределения для сумм независимых случайных величин, Гостехиздат, М., 1949.

8. Helstrom C. W., Resolution of Signals in White Gaussian Noise, *Proc. IRE,* **43** (Sept. 1955), 1111—1118.

9. Lebow I. L., McHugh P. G., Parker A. C., Rosen P., Wozencraft J. M., Application of Sequential Decoding to High Rate Data Communication on a Telephone Line, *IEEE Trans.,* **IT-9**, April 1963.

10. Massey J. L., Threshold Decoding, M. I. T. Press, Cambridge, Mass., 1963. (Русский перевод: Месси Дж., Пороговое декодирование, «Мир», М., 1966.)

11. Perry K. M., Wozencraft J. M., SECO: A Self Regulating Error Correcting Coder-Decoder, *IRE Trans.*, **IT-8**, 5 (Sept. 1962), 129—135.

12. Peterson W. W., Error-Correcting Codes, M. I. T. Press and Wiley, New York, 1961. (Русский перевод: Питерсон У., Коды, исправляющие ошибки, «Мир», М., 1964.)

13. Pierce J. R., Theoretical Diversity Improvement in Frequency Shift Keying, *Proc. IRE*, **46** (1958), 903—910.

14. Reiffen B., Sequential Decoding for Discrete Input Memoryless Channels, *IRE Trans.*, **IT-8**, 3 (April 1962), 208—220.

15. Shannon C. E., Certain Results in Coding Theory for Noisy Channels, *Inf. and Control.*, 1, 6—25 (1957). (Русский перевод: сб. Шеннон К., Работы по теории информации и кибернетике, ИЛ, М., 1963, 509.)

16. Wozencraft J. M., Horstein M., Coding for Two Way Channels, *Fourth London Symposium on Information Theory*, September, 1960.

17. Wozencraft J. M., Reiffen B., Sequential Decoding, Technology Press and Wiley, New York, 1961. (Русский перевод: Возенкрафт Дж., Рейффен Б., Последовательное декодирование, ИЛ, М., 1963.)

18. Zierler N., A Class of Cyclic Linear Error-Correcting Codes in p^m Symbols, M. I. T. Lincoln Laboratory, Group Report 55—19, Lexington, Mass., January, 1960.

ОГЛАВЛЕНИЕ

Р. Галлагер

КОДЫ С МАЛОЙ ПЛОТНОСТЬЮ ПРОВЕРОК НА ЧЕТНОСТЬ

Редактор *Н. Н. Щербиновская* Художник *Л. Г. Лорский*
Художественный редактор *В. И. Шаповалов*
Технический редактор *А. В. Грушин* Корректор *В. И. Бедель*

Сдано в производство 25/II 1966 г. Подписано к печати 24/IX 1966 г.
Бумага 84×108$^{1}/_{32}$=2,25 бум. л. 7,56 печ. л. Уч.-изд. л. 6,37.
Изд. № 1/3318. Цена 45 к. Зак. 104

ИЗДАТЕЛЬСТВО «М И Р»
Москва, 1-й Рижский пер., 2

Ленинградская типография № 2 имени Евгении Соколовой
Главполиграфпрома Комитета по печати при Совете Министров СССР,
Измайловский проспект, 29